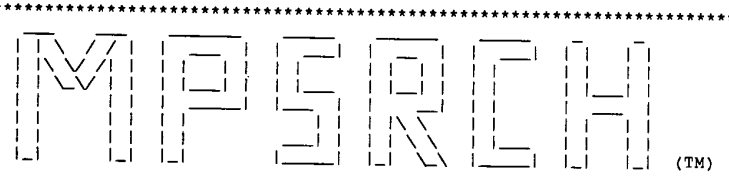
****



(TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:42:44 1999; MasPar time 465.47 Seconds
1393.481 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-7
Description: (51-284) from US09092296.seq
Perfect Score: 234
N.A. Sequence: 51 ATGGGGTCTGGGCTGCCCCT.....ACCATGTTGTCTGCAACACA 284
Comp: TACCCCAGACCCGACGGGGA.....TGGTACAACAGACGTTGTGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl58
1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi
Database: genbank111
17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 9.653; Variance 4.972; scale 1.941

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	92	39.3	47323	31	AC005937	Homo sapiens clone UWG	1.14e-47
	2	48	20.5	7218	25	I66494	Sequence 14 from paten	1.55e-15
	3	33	14.1	74371	31	AC005369	Homo sapiens chromosom	9.92e-06
	4	32	13.7	216021	31	HUAC004787	Homo sapiens Chromosom	4.05e-05
c	5	29	12.4	215	25	I28278	Sequence 5 from patent	2.49e-03
c	6	28	12.0	965	25	AR024229	Sequence 22 from paten	9.45e-03
	7	27	11.5	215	25	I28278	Sequence 5 from patent	3.51e-02
c	8	27	11.5	216021	31	HUAC004787	Homo sapiens Chromosom	3.51e-02
	9	25	10.7	565	25	E04076	gDNA encoding envelope	4.53e-01
c	10	25	10.7	6252	29	AB002302	Human mRNA for KIAA030	4.53e-01
c	11	24	10.3	60	25	A62989	Sequence 1 from Patent	1.56e+00
c	12	24	10.3	1056	23	MVU87256	Mustela vison GT dinuc	1.56e+00
	13	23	9.8	1056	23	MVU87256	Mustela vison GT dinuc	5.25e+00
c	14	23	9.8	1514	31	HSU47925	Human protein A altern	5.25e+00
	15	23	9.8	1678	32	MMHIS453	Mouse histone H4 gene.	5.25e+00
c	16	23	9.8	2277	25	AR004980	Sequence 2 from patent	5.25e+00
c	17	23	9.8	2516	31	HSU47928	Human protein A altern	5.25e+00
c	18	23	9.8	3032	31	AF104918	Homo sapiens myeloid/l	5.25e+00
	19	23	9.8	3290	29	HUMPFUCAS	H.sapiens fucosidase p	5.25e+00
c	20	23	9.8	46251	30	AD000671	Homo sapiens DNA from	5.25e+00
c	21	23	9.8	133457	30	AC003999	Human PAC clone DJ1139	5.25e+00
	22	23	9.8	167228	31	AC005552	Homo sapiens chromosom	5.25e+00
c	23	23	9.8	175902	20	AC006992	Homo sapiens clone NH0	5.25e+00
	24	23	9.8	203418	19	AC004947	Homo sapiens clone DJ1	5.25e+00
c	25	23	9.8	222930	30	HSU47924	Human chromosome 12p13	5.25e+00
c	26	22	9.4	30	25	A62994	Sequence 6 from Patent	1.71e+01
c	27	22	9.4	108	21	D87227	Trypanosoma cruzi mRNA	1.71e+01
	28	22	9.4	1004	27	SOYHRGPA	Soybean hydroxyproline	1.71e+01
c	29	22	9.4	1420	32	RATBAS	Rattus norvegicus beta	1.71e+01
c	30	22	9.4	1789	29	HSCYPINT1	H.sapiens CYP11B1 gene	1.71e+01
c	31	22	9.4	1945	24	CHKCONN45	Chicken connexin-45 mR	1.71e+01
c	32	22	9.4	2180	21	DROSYT	D.melanogaster synapto	1.71e+01
	33	22	9.4	2223	32	AF081783	Rattus norvegicus amil	1.71e+01
c	34	22	9.4	2460	32	AB010742	Mus musculus mRNA for	1.71e+01
	35	22	9.4	3043	32	AB010141S4	Mus musculus gene for	1.71e+01
	36	22	9.4	3247	32	AB010140	Mus musculus mRNA for	1.71e+01
	37	22	9.4	4434	28	GMU44838	Glycine max extensin (	1.71e+01
c	38	22	9.4	6910	29	HUMCYPBB	Human CYP11B2 gene for	1.71e+01
	39	22	9.4	67919	20	AC006542	WORKING DRAFT SEQUENCE	1.71e+01
c	40	22	9.4	124700	22	AC005558	Drosophila melanogaste	1.71e+01
	41	22	9.4	133457	30	AC003999	Human PAC clone DJ1139	1.71e+01
	42	22	9.4	167217	30	AC002350	Homo sapiens 12q24 PAC	1.71e+01
c	43	22	9.4	175339	31	AC005772	Homo sapiens chromosom	1.71e+01
	44	22	9.4	214646	20	AC006512	Homo sapiens clone RPC	1.71e+01
c	45	22	9.4	241407	20	AC003059	Mouse Chromosome 10 BA	1.71e+01

ALIGNMENTS

RESULT 1
LOCUS AC005937 47323 bp DNA PRI 05-NOV-1998
DEFINITION Homo sapiens clone UWGC:370M23.002 from 6p21, complete sequence.
ACCESSION AC005937
NID g3845393
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 47323)
AUTHORS Janer,M., Guillaudeux,T., Vu,Q., Kutyavin,T., Harter,H. and Geraghty,D.E.
TITLE Large scale sequence analysis of the human MHC class I region
JOURNAL Unpublished (1998)
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
REFERENCE 2 (bases 1 to 47323)
AUTHORS Geraghty,D.E. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhcrc.org)
COMMENT Overlapping Sequences:
5': UWGC:370m23.013 (Genbank Accession: AC005530)
3': UWGC:y67c112 (Genbank Accession: AC004211)

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than

1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 75.5%
DS or two chemistry coverage: 98.9%
Single stranded regions: 3

Sequence Validation:
This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BglII		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
1069.11	1050.00	889.55	866.00	30541.40	30653.00
20320.67	20855.00	1050.18	1015.00	3279.08	3231.00
2171.50	2147.00	7268.78	7196.00		
2560.20	2531.00	10085.80	9992.00		
4335.42	4269.00	11212.78	11131.00		
2698.62	2628.00				
1927.50	1887.00				
3130.46	3090.00				
2166.69	2129.00				
2044.67	2005.00				

```
FEATURES             Location/Qualifiers
     source          1. .47323
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21"
                     /sub_clone="UWGC:370M23.002"
                     /clone_lib="Research Genetics BAC Library"
     repeat_region   3647. .3932
                     /rpt_family="Alu"
     repeat_region   complement(4999. .5277)
                     /rpt_family="Alu"
     repeat_region   6285. .6572
                     /rpt_family="Alu"
     repeat_region   complement(6972. .7050)
                     /rpt_family="MLT1"
     repeat_region   7286. .7584
                     /rpt_family="Alu"
     repeat_region   complement(8164. .8609)
                     /rpt_family="Alu"
     repeat_region   complement(21287. .21895)
                     /rpt_family="Alu"
     repeat_region   22715. .22957
                     /rpt_family="Alu"
     repeat_region   25510. .25802
                     /rpt_family="Alu"
     repeat_region   27835. .28010
                     /rpt_family="MER20"
     repeat_region   31295. .31594
                     /rpt_family="Alu"
     repeat_region   33515. .33767
                     /rpt_family="Alu"
     repeat_region   34223. .34290
                     /rpt_family="MIR"
     repeat_region   37372. .37648
                     /rpt_family="Alu"
     repeat_region   38526. .38700
                     /rpt_family="MER3"
     repeat_region   39583. .40010
                     /rpt_family="Alu"
     repeat_region   40046. .40156
                     /rpt_family="Alu"
     repeat_region   43194. .43372
                     /rpt_family="MER5"
     variation       43325
                     /note="clonal variation with 3' overlapping clone"
     variation       44149
                     /note="clonal variation with 3' overlapping clone"
     variation       44451
                     /note="clonal variation with 3' overlapping clone"
     variation       44537
                     /note="clonal variation with 3' overlapping clone"
     variation       44814
                     /note="clonal variation with 3' overlapping clone"
     variation       44965
                     /note="clonal variation with 3' overlapping clone"
     variation       45760
                     /note="clonal variation with 3' overlapping clone"
     variation       45900
                     /note="clonal variation with 3' overlapping clone"
     variation       46851
                     /note="clonal variation with 3' overlapping clone"
     variation       46859
                     /note="clonal variation with 3' overlapping clone"
     variation       47032
                     /note="clonal variation with 3' overlapping clone"
     variation       47240. .47256
                     /note="clonal variation with 3' overlapping clone -
                     insertion of 17bp repeat"
BASE COUNT    11556 a  11489 c  12284 g  11994 t
ORIGIN

  Query Match           39.3%;  Score 92;  DB 31;  Length 47323;
  Best Local Similarity 100.0%;  Pred. No. 1.14e-47;
  Matches    92;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  35465 CAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCT 35524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    109 CAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCT 168

Db  35525 ATGAGTCCAGCTTCCTGGAATTGCTTGAAAAG 35556
          ||||||||||||||||||||||||||||||||
Qy    169 ATGAGTCCAGCTTCCTGGAATTGCTTGAAAAG 200


RESULT    2
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

```
  Query Match           20.5%;  Score 48;  DB 25;  Length 7218;
  Best Local Similarity   1.8%;  Pred. No. 1.55e-15;
  Matches     4;  Conservative  130;  Mismatches  86;  Indels   0;  Gaps   0;

Db   1221 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1280
          :: :::::: :::::::::: ::::::::  :  ::: : :    :    ::   : :
Qy     63 CTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATG 122

Db   1281 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1340
          :::: :  ::    ::        ::::::::  :   :::::::: :   :::  ::::
Qy    123 ACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTC 182

Db   1341 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1400
          ::    :: :::      :::: :::::::::: :::::::::     ::  : :: ::
Qy    183 CTGGAATTGCTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACC 242

Db   1401 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440
          :::: :: : :    ::::  : :: : :: :::| | ||
Qy    243 CTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAACA 282


RESULT    3
LOCUS       AC005369    74371 bp    DNA             PRI       01-AUG-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 119j3 (LBNL H175), complete
            sequence.
ACCESSION   AC005369
NID         g3367505
VERSION     AC005369.1  GI:3367505
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 74371)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 74371)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 74371)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .74371
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="119j3"
                     /chromosome="5"
                     /note="LBNL H175"
     repeat_region   893. .1030
                     /rpt_family="Alu"
     repeat_region   2295. .2438
                     /rpt_family="Alu"
     repeat_region   2818. .2859
                     /note="(GT)21"
                     /rpt_type=tandem
                     /rpt_unit=GT
     misc_feature    join(3246. .3410,3721. .3828)
                     /standard_name="RLF"
                     /note="65% & 69% protein identity GenPept:U22377"
     repeat_region   3431. .3724
                     /rpt_family="Alu"
     repeat_region   3707. .3728
                     /note="(A)22"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   4366. .4661
                     /rpt_family="Alu"
     repeat_region   5327. .5602
                     /rpt_family="Alu"
     repeat_region   6586. .6956
                     /rpt_family="L1"
     repeat_region   6647. .6684
                     /note="(CA)19"
                     /rpt_type=tandem
                     /rpt_unit=CA
     repeat_region   7113. .7373
                     /rpt_family="Alu"
     repeat_region   complement(7830. .8185)
                     /standard_name="possible repeat"
     repeat_region   8258. .8503
                     /rpt_family="Alu"
     repeat_region   9070. .9387
                     /rpt_family="Alu"
     repeat_region   complement(9740. .9845)
                     /rpt_family="MER42"
     repeat_region   complement(10440. .11015)
                     /rpt_family="Alu"
     repeat_region   11950. .12250
                     /rpt_family="Alu"
     repeat_region   12057. .12085
                     /note="(A)29"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   12365. .12645
                     /rpt_family="Alu"
     repeat_region   13727. .13750
                     /note="(AC)12"
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   13783. .14024
                     /rpt_family="L1"
     repeat_region   14175. .14470
                     /rpt_family="Alu"
     repeat_region   complement(14906. .15259)
                     /standard_name="possible repeat"
     repeat_region   15300. .15613
                     /rpt_family="Alu"
     repeat_region   16671. .16690
                     /note="(T)20"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(16675. .16977)
                     /rpt_family="Alu"
     repeat_region   complement(16993. .17085)
                     /rpt_family="MER42"
     repeat_region   complement(17678. .18276)
                     /rpt_family="Alu"
     repeat_region   19305. .19583
                     /rpt_family="Alu"
     repeat_region   19914. .19945
                     /note="(GTTT)8"
                     /rpt_type=tandem
                     /rpt_unit=GTTT
     repeat_region   complement(19943. .20222)
                     /rpt_family="Alu"
     repeat_region   21202. .21496
                     /rpt_family="Alu"
     repeat_region   21736. .22035
                     /rpt_family="Alu"
     repeat_region   22017. .22038
                     /note="(A)22"
                     /rpt_type=tandem
```

```
                /rpt_unit=A
repeat_region   22321. .22457
                /rpt_family="Alu"
repeat_region   complement(22638. .22981)
                /standard_name="possible repeat"
repeat_region   23022. .23326
                /rpt_family="Alu"
repeat_region   23473. .23761
                /rpt_family="Alu"
repeat_region   23744. .23767
                /note="(A)24"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   complement(24639. .24694)
                /rpt_family="MER42"
repeat_region   complement(25349. .25713)
                /rpt_family="Alu"
repeat_region   complement(25727. .26471)
                /rpt_family="Alu"
repeat_region   27191. .27477
                /rpt_family="Alu"
repeat_region   27774. .28057
                /rpt_family="Alu"
repeat_region   28040. .28066
                /note="(A)27"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   28712. .28930
                /rpt_family="MER20"
misc_feature    complement(28769. .28838)
                /note="GRAIL 2 excellent exon, frame 0"
repeat_region   28987. .29214
                /rpt_family="Alu"
misc_feature    join(29485. .29559,29598. .29779)
                /note="95% identity dbSTS:G14522 (SHGC-11312)"
misc_feature    29495. .29976
                /standard_name="AI027942"
                /note="100% identity EST ov84a10.x1"
misc_feature    complement(30401. .30536)
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   complement(30682. .30733)
                /rpt_family="MIR"
misc_feature    complement(31573. .31724)
                /note="GRAIL 2 excellent exon, frame 1"
misc_feature    complement(32159. .32232)
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   32388. .32488
                /rpt_family="MLT1"
repeat_region   32617. .32908
                /rpt_family="Alu"
repeat_region   32977. .33088
                /rpt_family="MLT1"
repeat_region   complement(33670. .33785)
                /rpt_family="Alu"
misc_feature    complement(34021. .34144)
                /note="GRAIL 2 excellent exon, frame 2"
misc_feature    complement(35238. .35331)
                /note="GRAIL 2 excellent exon, frame 1"
repeat_region   36392. .36663
                /rpt_family="Alu"
misc_feature    36901. .37222
                /note="100% identity EST ou55c09.x1"
                /db_xref="dbEST:AI025011"
STS             36901. .37164
                /standard_name="STSG-9983"
                /db_xref="dbSTS:G26554"
misc_feature    complement(37595. .37654)
                /note="GRAIL 2 excellent exon, frame 0"
misc_feature    complement(join(38063. .38218,38462. .38578,38741. .38995,
                39071. .39205,39532. .39630,39935. .40048,40300. .40410,
                40503. .40661,41868. .41972,42103. .42225,42492. .42569,
                44379. .44507))
                /standard_name="histidyl-tRNA synthetase"
                /note="78%-100% protein identity GenPept:U18937"
misc_feature    complement(38069. .38215)
                /note="GRAIL 2 excellent exon, frame 0"
...
Note: remainder of annotations omitted.

  Query Match           14.1%;  Score 33;  DB 31;  Length 74371;
  Best Local Similarity  17.9%;  Pred. No. 9.92e-06;
  Matches    19;  Conservative    53;  Mismatches  33;  Indels    1;  Gaps    1;

Db  15901 TTTTKWGASRSRGWSTCKCKCTSTSKCSCMSRSKSKRGWGYRSWKKYRCAMWMTCKSSKC 15960
          |||| :||:: : ::   |: |: ::|:  :   :::|: : : :: : |: :||::: 
Qy    149 TTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAAAAGCTCTGCCT 208

Db  15961 WCWSYRMRMKCYSCSYCYCSSGKKYWCRCSMYWYTCYYSYKYYWSM 16006
            |::: : ::|::|:: : ::|  :: :|: ::: |:::  : :::
Qy    209 CCTCCTCCATCTCCCTTCAGGGACCA-GCGTCACCCTCCACCATGC 253


RESULT    4
LOCUS       HUAC004787 216021 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
NID         g3337381
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES            Location/Qualifiers
     source         1. .216021
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="16"
                    /map="#16q21-22"
                    /clone="A-952F10"
     STS            27765. .27872
                    /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                    sapiens"
                    /db_xref="dbSTS:G02280"
     STS            73826. .73943
```

```
                /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                sapiens"
                /db_xref="dbSTS:G02122"
     STS        175801. .175945
                /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                /db_xref="dbSTS:G09703"
     STS        175810. .175945
                /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                /db_xref="dbSTS:G09935"
     STS        199463. .199572
                /note="9824, WI-3555, Chr. 16, Homo sapiens"
                /db_xref="dbSTS:G04338"
BASE COUNT    60960 a  51778 c  49172 g  53987 t    124 others
ORIGIN

  Query Match           13.7%;  Score 32;  DB 31;  Length 216021;
  Best Local Similarity  12.0%;  Pred. No. 4.05e-05;
  Matches   11;  Conservative   51;  Mismatches  30;  Indels   0;  Gaps   0;

Db   1457 KKYMWKSMRRARRRGSAGKKKKKYYYYYCYYYYYYYYCMGRAMMAAAWYYKRRSCMAWYY 1516
           :: ::::::::|:        :  :::::: ::::::::|:|: :: |   : : :|    :
Qy    189 TTGCTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCTCCAC 248

Db   1517 YMRGRARKTYYYMARRRGCARSYKKYMAMYWA 1548
           :: | |:: ::::|: : || ::::::  ::|
Qy    249 CATGCAAGATCTCAACACCATGTTGTCTGCAA 280


RESULT    5
LOCUS       I28278        215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
VERSION     I28278.1  GI:1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="unknown"
BASE COUNT       15 a       8 c      25 g      26 t    141 others
ORIGIN

  Query Match           12.4%;  Score 29;  DB 25;  Length 215;
  Best Local Similarity  15.1%;  Pred. No. 2.49e-03;
  Matches   26;  Conservative   69;  Mismatches  75;  Indels   2;  Gaps   2;

Db      1 MTNVTMSSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNK 60
          :| :|::: ::::: |:  ::|:::|   :::  :|   |  :   :||: :  : |: :
Cp    235 CTGGTCCCTGAAGGGAGATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTG 176

Db     61 YSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSY 120
           :  : : |  :|||:|: : :   :   ::: :|::       : |  :   :  :::
Cp    175 GAC-TCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACC 117

Db    121 -GSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTKKHAKNSADGKVGS 171
           |::    |   :||: |   : ::  |: :: ::   ::: : :|:|::|:
Cp    116 CGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGGC 65


RESULT    6
LOCUS       AR024229      965 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 22 from patent US 5795961.
ACCESSION   AR024229
NID         g3977523
VERSION     AR024229.1  GI:3977523
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and
            Kitamura,K.
  TITLE     Recombinant human anti-Lewis b antibodies
  JOURNAL   Patent: US 5795961-A 22 18-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .965
                     /organism="unknown"
BASE COUNT      192 a    170 c    226 g    205 t    172 others
ORIGIN

  Query Match           12.0%;  Score 28;  DB 25;  Length 965;
  Best Local Similarity  16.3%;  Pred. No. 9.45e-03;
  Matches   26;  Conservative   65;  Mismatches  69;  Indels   0;  Gaps   0;

Db    771 AARTGGGTKKGTRHTVHVSGGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKR 830
          ||: | | ::| :  : ::| ::  :| | ::     :  ::|: :| :|| :|    ::
Cp    225 AAGGGAGATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGG 166

Db    831 GRVTMADTSSNSRSSVTAADTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHTVKDMTSSSS 890
          :: : :|:   :  :  | | :::|::   :::: |  : || :|:::: :  :|:  :
Cp    165 AGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTC 106

Db    891 ASVGDRVTTCRSSTTHGNGNTYYWYKGKAKYRVSNRSGVS 930
            |:::  :::  | |  ::: : :  ::: ::|::
Cp    105 CATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 66


RESULT    7
LOCUS       I28278        215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
VERSION     I28278.1  GI:1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="unknown"
BASE COUNT       15 a       8 c      25 g      26 t    141 others
ORIGIN

  Query Match           11.5%;  Score 27;  DB 25;  Length 215;
  Best Local Similarity  20.0%;  Pred. No. 3.51e-02;
  Matches   22;  Conservative   42;  Mismatches  45;  Indels   1;  Gaps   1;

Db     18 CNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAK 77
          | :: : :| |  ::| :   :  |:|: :| ::: : |:  : :|   :|   :|| :
Qy     89 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG-AAGGAGT 147

Db     78 THYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
          :::|:| :: |: : :| ::  :   :: | |:   :||: :|: : |:
Qy    148 CTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAA 197


RESULT    8
LOCUS       HUAC004787 216021 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
NID         g3337381
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES              Location/Qualifiers
     source           1. .216021
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="16"
                      /map="#16q21-22"
                      /clone="A-952F10"
     STS              27765. .27872
                      /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                      sapiens"
                      /db_xref="dbSTS:G02280"
     STS              73826. .73943
                      /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                      sapiens"
                      /db_xref="dbSTS:G02122"
     STS              175801. .175945
                      /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                      /db_xref="dbSTS:G09703"
     STS              175810. .175945
                      /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                      /db_xref="dbSTS:G09935"
     STS              199463. .199572
                      /note="9824, WI-3555, Chr. 16, Homo sapiens"
                      /db_xref="dbSTS:G04338"
BASE COUNT    60960 a  51778 c  49172 g  53987 t    124 others
ORIGIN

  Query Match           11.5%;  Score 27;  DB 31;  Length 216021;
  Best Local Similarity  13.8%;  Pred. No. 3.51e-02;
  Matches    18;  Conservative   63;  Mismatches  48;  Indels   1;  Gaps   1;

Db   1457 KKYMWKSMRRARRRGSAGKKKKKYYYYYCYYYYYYYYCMGRAMMAAAWYYKRRSCMAWYY 1516
          :: : ::::: :::|:||: ::     : ::::   : :   ::  |: :::: |  ::
Cp    227 TGAAGGGAGATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTGGACTCATA 168

Db   1517 YMRGRARKTYYYMARRRGCARSYK-KYMAMYWAAWTTWCAAAAMAAMYTTYYWMMWYYKK 1575
             :|:|::|: ::|:::   :::  :::| ::  : |:  |||   ::  ::  : ':::::
Cp    167 GGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGT 108

Db   1576 WMYTCCTGAG 1585
           :::  ||||
Cp    107 TCCATGTGAG 98


RESULT    9
LOCUS       E04076        565 bp    RNA             PAT       29-SEP-1997
DEFINITION  gDNA encoding envelope region of type C hepatitis virus.
ACCESSION   E04076
NID         g2172286
VERSION     E04076.1  GI:2172286
KEYWORDS    JP 1992349885-A/1.
SOURCE      Hepatitis C virus.
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepatitis C-like viruses.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Morinaga,T., Chayama,K., Kumada,H. and Ichikawa,Y. .
  TITLE     NUCLEIC ACID FRAGMENT WITHIN ENVELOPE REGION OF HEPATITIS C VIRUS
            AND METHOD FOR DETECTING THE SAME
  JOURNAL   Patent: JP 1992349885-A 1 04-DEC-1992;
            TEIJIN LTD
COMMENT     OS   Hepatitis C virus
            PN   JP 1992349885-A/1
            PD   04-DEC-1992
            PF   29-MAY-1991 JP 1991152169
            PI   MORINAGA TSUTAE, CHAYAMA KAZUAKI, KUMADA HIROMITSU, PI
            ICHIKAWA YATARO
            PC   C12N15/10,C12Q1/68,C12Q1/70//C12N15/11;
            CC   strandedness: Single;
            CC   topology: Linear;.
FEATURES              Location/Qualifiers
     source           1. .565
                      /organism="Hepatitis C virus"
                      /db_xref="taxon:11103"
BASE COUNT       60 a     93 c    107 g     85 t    220 others
ORIGIN

  Query Match           10.7%;  Score 25;  DB 25;  Length 565;
  Best Local Similarity  34.1%;  Pred. No. 4.53e-01;
  Matches    30;  Conservative   29;  Mismatches  28;  Indels   1;  Gaps   1;

Db    400 GGRTCYGTYTTYCTYRTYTCYCABYTGTTTACCYTCTMBMCYMGSMDGYAYGDGACRDYR 459
          ||:||:|   | | : |:  :| ::| || |||:||  :  ::|:   :|:|: ||:: :
Qy     54 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 113

Db    460 CARGRYTGYAAYTGYTCDMTYTAYSCYG 487
          | :|::  : |:| : |:: : | :|:|
Qy    114 CCGGGTA-TGACTTTGCAACTGAAGCTG 140


RESULT   10
LOCUS       AB002302     6252 bp    mRNA            PRI       13-FEB-1999
DEFINITION  Human mRNA for KIAA0304 gene, complete cds.
ACCESSION   AB002302
NID         g2224548
VERSION     AB002302.1  GI:2224548
KEYWORDS    KIAA0304.
SOURCE      Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK
            plus clone:HG0016.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6252)
  AUTHORS   Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930,
            Fax:+81-438-52-3931)
REFERENCE   2  (sites)
```

```
  AUTHORS   Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
  JOURNAL   DNA Res. 4 (2), 141-150 (1997)
  MEDLINE   97349984
FEATURES             Location/Qualifiers
     source          1. .6252
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HG0016"
                     /clone_lib="pBluescriptII SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1343. .5932
                     /gene="KIAA0304"
     CDS             1343. .5932
                     /gene="KIAA0304"
                     /codon_start=1
                     /protein_id="BAA20763.1"
                     /db_xref="PID:d1021600"
                     /db_xref="PID:g2224549"
                     /db_xref="GI:2224549"
                     /translation="MGGLSVLTSVPGGPPMVCLLCASKGLHELVFCQVCCDPFHPFCL
                     EEAERPLPQHHDTWCCRRCKFCHVCGRKGRGSKHLLECERCRHAYHPACLGPSYPTRA
                     TRKRRHWICSACVRCKSCGATPGKNWDVEWSGDYSLCPRCTQLYEKGNYCPICTRCYE
                     DNDYESKMMQCAQCDHWVHAKCEGLSDEDYEILSGLPDSVLYTCGPCAGAAQPRWREA
                     LSGALQGGLRQVLQGLLSSKVVGPLLLCTQCGPDGKQLHPGPCGLQAVSQRFEDGHYK
                     SVHSFMEDMVGILMRHSEEGETPDRRAGGQMKGLLLKLLESAFGWFDAHDPKYWRRST
                     RLPNGVLPNAVLPPSLDHVYAQWRQQEPETPESGQPPGDPSAAFQGKDPAAFSHLEDP
                     RQCALCLKYGDADSKEAGRLLYIGQNEWTHVNCAIWSAEVFEENDGSLKNVHAAVARG
                     RQMRCELCLKPGATVGCCLSSCLSNFHFMCARASYCIFQDDKKVFCQKHTDLLDGKEI
                     VNPDGFDVLRRVYVDFEGINFKRKFLTGLEPDAINVLIGSIRIDSLGTLSDLSDCEGR
                     LFPIGYQCSRLYWSTVDARRRCWYRCRILEYRPWGPREEPAHLEAAEENQTIVHSPAP
                     SSEPPGGEDPPLDTDVLVPGAPERHSPIQNLDPPLRPDSGSAPPPAPRSFSGARIKVP
                     NYSPSRRPLGGVSFGPLPSPGSPSSLTHHIPTVGDPDFPAPPRRSRRPSPLAPRPPPS
                     RWASPPLKTSPQLRVPPPTSVVTALTPTSGELAPPGPAPSPPPPEDLGPDFEDMEVVS
                     GLSAADLDFAASLLGTEPFQEEIVAAGAMGSSHGGPGDSSEEESSPTSRYIHFPVTVV
                     SAPGLAPSATPGAPRIEQLDGVDDGTDSEAEAVQQPRGQGTPPSGPGVVRAGVLGAAG
                     DRARPPEDLPSEIVDFVLKNLGGPGDGGAGPREESLPPAPPLANGSQPSQGLTASPAD
                     PTRTFAWLPGAPGVRVLSLGPAPEPPKPATSKIILVNKLGQVFVKMAGEGEPVPPPVK
                     QPPLPPTISPTAPTSWTLPPGPLLGVLPVVGVVRPAPPPPPPPLTLVLSSGPASPPRQ
                     AIRVKRVSTFSGRSPPAPPPYKAPRLDEDGEASEDTPQVPGLGSGGFSRVRMKTPTVR
                     GVLDLDRPGEPAGEESPGPLQERSPLLPLPEDGPPQVPDGPPDLLLESQWHHYSGEAS
                     SSEEEPPSPDDKENQAPKRTGPHLRFEISSEDGFSVEAESLEGAWRTLIEKVQEARGH
                     ARLRHLSFSGMSGARLLGIHHDAVIFLAEQLPGAQRCQHYKFRYHQQGEGQEEPPLNP
                     HGAARAEVYLRKCTFDMFNFLASQHRVLPEGATCDEEEDEVQLRSTRRATSLELPMAM
                     RFRHLKKTSKEAVGVYRSAIHGRGLFCKRNIDAGEMVIEYSGIVIRSVLTDKREKFYD
                     GKGIGCYMFRMDDFDVVDATMHGNAARFINHSCEPNCFSRVIHVEGQKHIVIFALRRI
                     LRGEELTYDYKFPIEDASNKLPCNCGAKRCRRFLN"
BASE COUNT     1160 a    2028 c    1860 g    1204 t
ORIGIN

  Query Match           10.7%;  Score 25;  DB 29;  Length 6252;
  Best Local Similarity  78.8%;  Pred. No. 4.53e-01;
  Matches   41;  Conservative    0;  Mismatches  10;  Indels    1;  Gaps    1;

Db       8 TGCCGCTGAGCCCTGGAGGGCAGATGGAGGAGGTGGCCGGGGCTGTCAAGCA 59
          || |||||  ||||| |||| |||||||||||| ||| | |  | |||||||
Cp     240 TGACGCTGGTCCCTGAAGGG-AGATGGAGGAGGAGGCAGAGCTTTTCAAGCA 190


RESULT   11
LOCUS       A62989         60 bp    DNA              PAT       12-MAR-1998
DEFINITION  Sequence 1 from Patent WO9720068.
ACCESSION   A62989
NID         g3716861
VERSION     A62989.1  GI:3716861
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 60)
  AUTHORS   Oerum,H. and Seeger,C.
  TITLE     METHOD FOR GENERATING MULTIPLE DOUBLE STRANDED NUCLEIC ACIDS
  JOURNAL   Patent: WO 9720068-A 1 05-JUN-1997;
            BOEHRINGER MANNHEIM GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        0 a      30 c       0 g       0 t      30 others
ORIGIN

  Query Match           10.3%;  Score 24;  DB 25;  Length 60;
  Best Local Similarity  10.0%;  Pred. No. 1.56e+00;
  Matches     3;  Conservative   24;  Mismatches   3;  Indels    0;  Gaps    0;

Db      27 CCCCRRRRRRRRRRRRRRRRRRRRRRRR 56
           ||| ::::::::: :::::::::::: ::::
Cp     230 CCCTGAAGGGAGATGGAGGAGGAGGCAGAG 201


RESULT   12
LOCUS       MVU87256     1056 bp    DNA              MAM       02-JAN-1999
DEFINITION  Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION   U87256
NID         g4099442
VERSION     U87256.1  GI:4099442
KEYWORDS    .
SOURCE      American mink.
  ORGANISM  Mustela vison
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
            Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /chromosome="1"
                     /map="1q"
                     /note="primers: 1167F: agcccctgcatatctacttctt, 1167R:
                     gaggatcttaccgctgttgag"
     primer_bind     98. .119
                     /standard_name="1167F"
     primer_bind     complement(300. .320)
                     /standard_name="1167R"
BASE COUNT      211 a     221 c     210 g     225 t     189 others
ORIGIN

  Query Match           10.3%;  Score 24;  DB 23;  Length 1056;
  Best Local Similarity  10.8%;  Pred. No. 1.56e+00;
  Matches     7;  Conservative   40;  Mismatches  17;  Indels    1;  Gaps    1;

Db     611 KRMRRCSRYMMMRMCVGSGTWARCCCCDKKSGSGDKSHCKSKRKYKWMDRYHBC-KSMCA 669
           ::::: : :: : :|:|:  :   || :::: : ::::  :::: ::::: ::  :::||
Cp     130 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 71

Db     670 MVRGK 674
           :::|:
Cp      70 AGGGG 66


RESULT   13
LOCUS       MVU87256     1056 bp    DNA              MAM       02-JAN-1999
DEFINITION  Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION   U87256
NID         g4099442
VERSION     U87256.1  GI:4099442
KEYWORDS    .
SOURCE      American mink.
```

```
  ORGANISM  Mustela vison
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
            Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /chromosome="1"
                     /map="1q"
                     /note="primers: 1167F: agcccctgcatatctacttctt, 1167R:
                     gaggatcttaccgctgttgag"
     primer_bind     98. .119
                     /standard_name="1167F"
     primer_bind     complement(300. .320)
                     /standard_name="1167R"
BASE COUNT      211 a    221 c    210 g    225 t    189 others
ORIGIN

  Query Match            9.8%;  Score 23;  DB 23;  Length 1056;
  Best Local Similarity  14.8%;  Pred. No. 5.25e+00;
  Matches    16;  Conservative   52;  Mismatches  39;  Indels   1;  Gaps   1;

Db    644 GDKSHCKSKRKYKWMDRYHBCKSMCAMVRGKWSAGTGCMARGWRTWDSSCTSTKGARRRK 703
          |::::|:::: :: :  :::  : |  :  :::  | |  :| :  :: |   : | :::
Qy     53 GGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 112

Db    704 GMYARRKMYGTGAAWASATAYSTGAGTTBWRVTSWKMBKYKKTGMMHH 751
          |:: :::::|    : :|  ::  || |::::   :::::: ||:::
Qy    113 GCCGGGTATGACTTTGCAACTGA-AGCTGAAGGAGTCTTTTCTGACAA 159


RESULT   14
LOCUS       HSU47925     1514 bp    mRNA            PRI       09-MAY-1996
DEFINITION  Human protein A alternatively spliced form 1 (A-1) mRNA, complete
            cds.
ACCESSION   U47925
NID         g1208739
VERSION     U47925.1  GI:1208739
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Ansari-Lari,M.A., Muzny,D.M., Lu,J., Lu,F., Lilley,C.E., Spanos,S.,
            Malley,T. and Gibbs,R.A.
  TITLE     A gene-rich cluster between the CD4 and triosephosphate isomerase
            genes at human chromosome 12p13
  JOURNAL   Genome Res. 6 (4), 314-326 (1996)
  MEDLINE   96303695
REFERENCE   2  (bases 1 to 1514)
  AUTHORS   Ansari-Lari,M.A., Muzny,D.M. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-1996) M. Ali Ansari-Lari, Molecular and Human
            Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX
            77030, USA
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="Homo sapiens"
                     /note="sequence obtained from cDNA clone HIBBS19 dbEST
                     Accession Number T09310"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12p13"
                     /clone="HIBBS19"
     gene            1. .1514
                     /gene="A-1"
     exon            1. .104
                     /gene="A-1"
                     /number=1
     exon            105. .400
                     /gene="A-1"
                     /number=2
     CDS             386. .1300
                     /gene="A-1"
                     /note="alternatively spliced form 1 of three possible
                     forms"
                     /codon_start=1
                     /product="protein A-1"
                     /protein_id="AAC50463.1"
                     /db_xref="PID:g1208740"
                     /db_xref="GI:1208740"
                     /translation="MLSTGVVSFFSLKSDSAPPWMVLAVLWCSMAQTLLLPSFIWSCE
                     RYRADVRTVWEQCVAIMSEEDGDDDGGCDDYAEGRVCKVRFDANGATGPGSRDPAQVK
                     LLPGRHMLFPPLERVHYLQVPLSRRLSHDETNIFSTPREPGSFLHKWSSSDDIRVLPA
                     QSRALGGPPEYLGQRHRLEDEEDEEEAEGGGLASLRQFLESGVLGSGGGPPRGPGFFR
                     EEITTFIDETPLPSPTASPGHSPRRPRPLGLSPRRLSLGSPESRAVGLPLGLSAGRRC
                     SLTGGEESARAWGGSWGPGNPIFPQLTL"
     exon            401. .590
                     /gene="A-1"
                     /number=3
     exon            591. .748
                     /gene="A-1"
                     /number=4
     exon            749. .1514
                     /gene="A-1"
                     /number=5
     polyA_signal    1490. .1495
                     /gene="A-1"
     polyA_site      1514
                     /gene="A-1"
BASE COUNT      293 a    466 c    471 g    284 t
ORIGIN

  Query Match            9.8%;  Score 23;  DB 31;  Length 1514;
  Best Local Similarity  74.5%;  Pred. No. 5.25e+00;
  Matches    35;  Conservative    0;  Mismatches  12;  Indels   0;  Gaps   0;

Db   1201 AGGGAGACGCTGCTCCCTGACGGGGGGTGAAGAAAGTGCAAGGGCTT 1247
          |||| ||||||| ||||||| ||| | || || | | |  || ||||
Cp    244 AGGGTGACGCTGGTCCCTGAAGGGAGATGGAGGAGGAGGCAGAGCTT 198


RESULT   15
LOCUS       MMHIS453     1678 bp    DNA             ROD       12-SEP-1993
DEFINITION  Mouse histone H4 gene.
ACCESSION   X13236
NID         g51310
VERSION     X13236.1  GI:51310
KEYWORDS    histone; histone H4.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1678)
  AUTHORS   Schuemperli,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-1988) Schuemperli D., Zoologisches Institut, Abt.
            Entwicklungsbiologie, Baltzerstr. 4, CH-3012 Bern, Schweiz
REFERENCE   2  (bases 342 to 1241)
  AUTHORS   Meier,V.S., Bohni,R. and Schumperli,D.
  TITLE     Nucleotide sequence of two mouse histone H4 genes
  JOURNAL   Nucleic Acids Res. 17 (2), 795 (1989)
  MEDLINE   89128447
COMMENT     Nucleotides 242 - 341 and 1242 - 1430 have not been determined.
FEATURES             Location/Qualifiers
     source          1. .1678
                     /organism="Mus musculus"
                     /strain="BALB.A2g-Mx"
                     /db_xref="taxon:10090"
                     /clone_lib="lambda EMBL3."
```

```
                     /clone="53"
     promoter        557. .563
                     /note="pot. Sp1 binding site"
     misc_feature    570. .587
                     /note="H4 specific box"
     promoter        597. .604
                     /note="TATA-box"
     precursor_RNA   627. .1025
                     /note="pot. primary transcript"
     precursor_RNA   628. .1025
                     /note="pot. alt. primary transcript"
     CDS             660. .971
                     /note="histone H4 (AA 1 - 103)"
                     /codon_start=1
                     /protein_id="CAA31622.1"
                     /db_xref="PID:g51311"
                     /db_xref="GI:51311"
                     /db_xref="MGD:MGI:96099"
                     /db_xref="SWISS-PROT:P02304"
                     /translation="MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGV
                     KRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYRLKRQGRTLYGFG
                     G"
BASE COUNT      414 a    340 c    284 g    349 t    291 others
ORIGIN

  Query Match            9.8%;  Score 23;  DB 32;  Length 1678;
  Best Local Similarity  78.0%;  Pred. No. 5.25e+00;
  Matches    32;  Conservative    0;  Mismatches   9;  Indels   0;  Gaps   0;

Db   1028 GAATTCCTTAAAAAGCTGTGCATTTTGCTCCCTCACCCTTC 1068
           ||||| ||| ||||||| ||| |  | |||| || ||||||
Qy    186 GAATTGCTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTC 226


Search completed: Sun Oct 24 17:50:39 1999
Job time : 475 secs.
```

****

MPSRCH (TM)

****

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:59:00 1999; MasPar time 73.17 Seconds
685.248 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-7
Description: (51-284) from US09092296.seq
Perfect Score: 234
N.A. Sequence: 51 ATGGGGTCTGGGCTGCCCCT.....ACCATGTTGTCTGCAACACA 284
Comp: TACCCCAGACCCGACGGGGA.....TGGTACAACAGACGTTGTGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 7.815; Variance 4.996; scale 1.564

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	234	100.0	439	60	V84366	Human stomach carcino	5.37e-133
	2	40	17.1	91	9	Q51746	Oligonucleotide probe	9.33e-10
c	3	37	15.8	204	1	N81164	Base substituted E.co	4.31e-08
c	4	35	15.0	91	9	Q51746	Oligonucleotide probe	5.35e-07
	5	33	14.1	204	1	N81164	Base substituted E.co	6.41e-06
	6	29	12.4	114	12	Q70467	Generic DNA sequence	8.13e-04
	7	29	12.4	114	12	Q70470	Generic DNA sequence	8.13e-04
	8	28	12.0	114	12	Q70468	Generic DNA sequence	2.65e-03
	9	28	12.0	114	12	Q70472	Generic DNA sequence	2.65e-03
	10	28	12.0	114	12	Q70465	Generic DNA sequence	2.65e-03
	11	28	12.0	178	32	T76405	Human endothelin-1 an	2.65e-03
	12	27	11.5	114	12	Q70466	Generic DNA sequence	8.51e-03
	13	26	11.1	114	12	Q70469	Generic DNA sequence	2.70e-02
c	14	25	10.7	39	7	Q51787	Mixed oligonucleotide	8.42e-02
c	15	25	10.7	114	12	Q70470	Generic DNA sequence	8.42e-02
c	16	24	10.3	91	46	V44650	Mammalian DNA replica	2.58e-01
	17	24	10.3	114	12	Q70471	Generic DNA sequence	2.58e-01
c	18	24	10.3	114	12	Q70467	Generic DNA sequence	2.58e-01
c	19	24	10.3	114	12	Q70469	Generic DNA sequence	2.58e-01
c	20	23	9.8	114	12	Q70468	Generic DNA sequence	7.79e-01
	21	23	9.8	565	6	Q35072	HCV envelope region n	7.79e-01
c	22	23	9.8	2277	42	V13834	Homo sapiens ambiguit	7.79e-01
c	23	23	9.8	2277	41	V05370	Human telomerase p105	7.79e-01
c	24	22	9.4	36	2	Q11195	Ballast Constituent c	2.30e+00
	25	22	9.4	75	21	T13612	DC43 TSAR library gen	2.30e+00
	26	22	9.4	82	21	T13610	DC43 TSAR library gen	2.30e+00
	27	22	9.4	114	12	Q70473	Generic DNA sequence	2.30e+00
c	28	22	9.4	114	12	Q70466	Generic DNA sequence	2.30e+00
c	29	22	9.4	114	12	Q70465	Generic DNA sequence	2.30e+00
c	30	22	9.4	114	12	Q70472	Generic DNA sequence	2.30e+00
	31	22	9.4	250	32	T76438	Substance P antisense	2.30e+00
	32	22	9.4	264	32	T76445	Substance P receptor	2.30e+00
c	33	22	9.4	501	3	N50026	Sequence encoding new	2.30e+00
c	34	22	9.4	501	3	N50025	Sequence encoding new	2.30e+00
	35	22	9.4	3871	2	N71302	HSV-1 gB and surround	2.30e+00
c	36	21	9.0	68	21	T13584	TSAR-9 library genera	6.66e+00
c	37	21	9.0	74	21	T13613	DC43 TSAR library gen	6.66e+00
	38	21	9.0	81	21	T13611	DC43 TSAR library gen	6.66e+00
	39	21	9.0	908	51	V63025	B. malayi ankyrin cDN	6.66e+00
c	40	21	9.0	908	52	V63315	Nucleotide nBmAnk908	6.66e+00
c	41	21	9.0	911	52	V63312	Nucleotide nDiAnk911	6.66e+00
c	42	21	9.0	911	51	V63012	D. immitis ankyrin nD	6.66e+00
	43	21	9.0	2539	55	V08822	Gene No. 12 encoding	6.66e+00
c	44	21	9.0	5235	51	V63022	D. immitis ankyrin nD	6.66e+00
c	45	21	9.0	5503	52	V63314	Full length nucleotid	6.66e+00

ALIGNMENTS

```
RESULT    1
ID   V84366 standard; cDNA to mRNA; 439 BP.
AC   V84366;
DT   30-MAR-1999  (first entry)
DE   Human stomach carcinoma cDNA clone HP10408.
KW   Transmembrane protein; HP10408; human; stomach cancer; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             75..311
FT                   /*tag=  a
FT                   /note= "cDNA comprising the coding region (minus
FT                          the stop codon) is claimed (Claim 3)"
PN   WO9855508-A2.
PD   10-DEC-1998.
PF   03-JUN-1998; J02445.
PR   03-JUN-1997; JP-144948.
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENTRE.
PI   Kato S, Sekine S, Yamaguchi T;
DR   WPI; 99-045730/04.
DR   P-PSDB; W88498.
PT   New human proteins containing transmembrane domains and their
PT   encoding sequences  - useful in the preparation of antibodies and
PT   large-scale protein production, gene diagnosis, and gene therapy
PS   Claim 4; Page 135; 178pp; English.
CC   This is the nucleotide sequence of cDNA clone HP10408, which
CC   includes a coding region (also claimed) for a novel human
CC   transmembrane protein (see W88498).  The clone was isolated from a
CC   stomach cancer cDNA library using a signal sequence detection
CC   method, and by protein synthesis by in vitro translation.  The
CC   encoded protein has a putative signal sequence and a putative
CC   internal transmembrane domain.  The invention provides nucleotide
CC   sequences (see V84359-76) coding for 18 transmembrane proteins
```

```
CC   (see W88491-508), vectors containing such polynucleotides, and
CC   eukaryotic cells containing the vectors.  The proteins can be
CC   used as antigens or as compositions in the preparation of
CC   antibodies against the proteins.  The polynucleotides can be used
CC   as probes for gene diagnosis, and as gene sources for gene therapy
CC   and large-scale production of proteins encoded by the cDNA.  The
CC   host cells are used for the detection of ligands corresponding to
CC   the expressed proteins, and the screening of low mol.wt. medicines.
SQ   Sequence   439 BP;    89 A;    137 C;    109 G;    104 T;

  Query Match          100.0%;  Score 234;  DB 60;  Length 439;
  Best Local Similarity 100.0%;  Pred. No. 5.37e-133;
  Matches   234;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     75 atggggtctgggctgccccttgtcctcctcttgaccctccttggcagctcacatggaaca 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     51 ATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACA 110

Db    135 gggccgggtatgactttgcaactgaagctgaaggagtcttttctgacaaattcctcctat 194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    111 GGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTAT 170

Db    195 gagtccagcttcctggaattgcttgaaaagctctgcctcctcctccatctcccttcaggg 254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    171 GAGTCCAGCTTCCTGGAATTGCTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGG 230

Db    255 accagcgtcaccctccaccatgcaagatctcaacaccatgttgtctgcaacaca 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    231 ACCAGCGTCACCCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAACACA 284


RESULT    2
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match          17.1%;  Score 40;  DB 9;  Length 91;
  Best Local Similarity  0.0%;  Pred. No. 9.33e-10;
  Matches    0;  Conservative   45;  Mismatches   5;  Indels   0;  Gaps   0;

Db     11 ssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsv 60
          :::::::::::: ::::::: :: : :::::: ::::::::::::::::::
Qy    205 GCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCTCCACCATGCA 254


RESULT    3
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match          15.8%;  Score 37;  DB 1;  Length 204;
  Best Local Similarity  7.9%;  Pred. No. 4.31e-08;
  Matches    8;  Conservative   54;  Mismatches  39;  Indels   0;  Gaps   0;

Db     86 ymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynv 145
          :::|   : ::: : :: :  : : | ::| : :::::   :  |::::::::: :::  :
Cp    172 TCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGC 113

Db    146 hnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
          : :    |  :  ::|: ::    :   :::::|::::::|
Cp    112 CCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGAC 72


RESULT    4
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match          15.0%;  Score 35;  DB 9;  Length 91;
  Best Local Similarity  0.0%;  Pred. No. 5.35e-07;
  Matches    0;  Conservative   41;  Mismatches   6;  Indels   0;  Gaps   0;
```

```
Db      11 ssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvy 57
           :::::::::::: ::::::: :::: :: :::::::::: :: :::::
Cp     183 GGAAGCTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGC 137


RESULT    5
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match           14.1%;  Score 33;  DB 1;  Length 204;
  Best Local Similarity  11.7%;  Pred. No. 6.41e-06;
  Matches     9;  Conservative   40;  Mismatches  28;  Indels   0;  Gaps   0;

Db     109 daaawyccyrrsvkydccynachhddhyvybbbvynvhnhnncncccbnnhvchnvhbnn 168
           :|||: | : :::::  |:  |:: ::::::: :  :: :    | |:  :: : :::
Qy     195 GAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCTCCACCATGCA 254

Db     169 hrnwayvrhdarrddvh 185
           :: : :::: |  ::::
Qy     255 AGATCTCAACACCATGT 271


RESULT    6
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;   0 A;   2 C;   2 G;   2 T;

  Query Match           12.4%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity  1.0%;  Pred. No. 8.13e-04;
  Matches     1;  Conservative   31;  Mismatches  71;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnn 62
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|  :
Qy      53 GGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 112

Db      63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 105
           :  :  :  :  :  :     :  :  :  :  :  :  :  :
Qy     113 GCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTG 155


RESULT    7
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
```

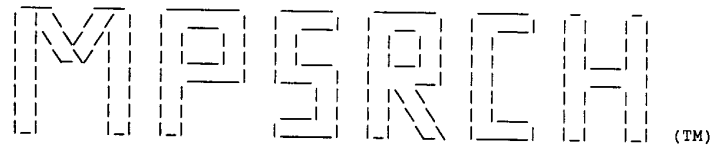
```
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    5 A;    10 C;    0 G;    0 T;

  Query Match           12.4%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   4.9%;  Pred. No. 8.13e-04;
  Matches     5;  Conservative   29;  Mismatches  69;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbcacnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :|    :  :  :  :| |  :  :  :  :  :  :  :     :
Qy     53 GGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 112

Db     63 bnnbnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnb 105
          :  :  :  :  :  :||   :  :  :  :  :  :  :  :
Qy    113 GCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTG 155


RESULT    8
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           12.0%;  Score 28;  DB 12;  Length 114;
  Best Local Similarity   2.7%;  Pred. No. 2.65e-03;
  Matches     3;  Conservative   31;  Mismatches  78;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :  :  :  :  :     :  :| |  :  :  :  :  :  :
Qy    134 GAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCT 193

Db     63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
          :     :  :  :  :  :  |  :  :  :     :  :  :  :  :  :
Qy    194 TGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCTC 245


RESULT    9
ID   Q70472 standard; DNA; 114 BP.
AC   Q70472;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58383.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70472 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)1(CAC)(NNB)11(CAC)(NNB)(CAC)(NNB)2Z(NNB)6
CC   -(CAC)(NNB)5(CAC)2(NNB)4. X and Y are flanking  restriction sites
CC   (X is not the same as Y) that are not specified further. The peptides
CC   generated by this and other generic sequences (Q70470-73) have invariant
CC   histidine residues incorporated into variant sequences. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The TSARs or compsns. comprising a TSAR binding domain can be used in
CC   vivo to deliver a chemically or biologically active moiety, eg. metal
CC   ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC   on the cell. They can also replace the function of macromolecules, eg.
```

```
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    6 A;    12 C;    0 G;    0 T;

  Query Match           12.0%;  Score 28;  DB 12;  Length 114;
  Best Local Similarity  11.3%;  Pred. No. 2.65e-03;
  Matches    12;  Conservative   24;  Mismatches  70;  Indels   0;  Gaps   0;

Db       9 bnnbnnbnnbnnbnnbnnbnnbnnbnnbcacnnbcacnnbnnbnnnnnnnnbnnbnn 68
          :  :  :  :     :  :     :  :  :| |  :| |  :  :        :  :
Qy     176 CAGCTTCCTGGAATTGCTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAG 235

Db      69 bnnbnnbnnbcacnnbnnbnnbnnbnnbcaccacnnbnnbnnbnnb 114
          :  :  :  :|||  :        :   |||||   :  :  :  :
Qy     236 CGTCACCCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAAC 281


RESULT   10
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           12.0%;  Score 28;  DB 12;  Length 114;
  Best Local Similarity   2.7%;  Pred. No. 2.65e-03;
  Matches     3;  Conservative   31;  Mismatches  78;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :  :|    :     :  :  :  :  :  :  :  :  :
Qy     134 GAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCT 193

Db      63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnb 114
          :     :  :  :  :  :  :  :  :  :  :     :  : ||  :  :  :
Qy     194 TGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCTC 245


RESULT   11
ID   T76405 standard; DNA; 178 BP.
AC   T76405;
DT   15-SEP-1997  (first entry)
DE   Human endothelin-1 antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Claim 5; Page 38; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human endothelin-1, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   178 BP;    0 A;    52 C;    46 G;    32 T;

  Query Match           12.0%;  Score 28;  DB 32;  Length 178;
  Best Local Similarity  25.0%;  Pred. No. 2.65e-03;
  Matches    18;  Conservative   32;  Mismatches  22;  Indels   0;  Gaps   0;

Db     100 cttggcbbgccbcbbbcbgcbgbgbgbbbbtcbtgbgcbbbtbbtccbttctgbbbbbbb 159
           || |||:: ||: :::|: |: : :| :::||:| :|| ::|:    :     :::::::
Qy      58 CTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGG 117

Db     160 gggbtcbbbbbc 171
           |  :  :::::|
Qy     118 GTATGACTTTGC 129


RESULT   12
ID   Q70466 standard; DNA; 114 BP.
AC   Q70466;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR-9 petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
```

```
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65152.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70466 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented as follows:X(NNB)1(TGC)(NNB)10(TGC)2(NNB)4Z(NNB)8(TGC)(NNB)
CC   -9Y. X and Y are flanking  restriction sites (X is not the same as Y)
CC   that are not specified further. Other generic sequences are shown in
CC   Q70466-68. Other specific peptides generated by these generic sequences
CC   are shown in R65151-54. TSARs are concatenated heterofunctional proteins
CC   or peptides, comprising at least two functional regions - a binding
CC   domain with affinity for a ligand and a second effector peptide portion
CC   that is chemically or biologically active.They may further comprise a
CC   linker peptide between the 2 domains.The oligonucleotides are also
CC   designed so that the expressed peptide contains 2 or 4 cysteine residues
CC   positioned in, or flanking, the unpredicted or variant residues. These
CC   residues confer some degree of conformational rigidity to the peptides.
CC   The TSARs or compsns. comprising a TSAR binding domain can be used in
CC   vivo to deliver a chemically or biologically active moiety, eg. metal
CC   ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC   on the cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match           11.5%;  Score 27;  DB 12;  Length 114;
  Best Local Similarity   4.9%;  Pred. No. 8.51e-03;
  Matches     5;  Conservative   28;  Mismatches  70;  Indels   0;  Gaps   0;

Db       9 bnnbnnbnnbnnbnnbnnbnnbnnbtgctgcnnbnnbnnbnnbnnnnnnnnbnnbnn 68
           :  :  :     :  :  :  :  :  :||| |  :     :  :        :  :
Qy     149 TTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAAAAGCTCTGCCT 208

Db      69 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnb 111
           :  :  :  :  :  :|   :  :  :  :  :  :  :  :
Qy     209 CCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCTCCACCAT 251


RESULT   13
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT       .           sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match           11.1%;  Score 26;  DB 12;  Length 114;
  Best Local Similarity   1.9%;  Pred. No. 2.70e-02;
  Matches     2;  Conservative   30;  Mismatches  74;  Indels   0;  Gaps   0;

Db       6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnnbnn 65
           :  :  :  :  :  :  :  :  :     :| |  :  :  :  :  :  :  :        :
Qy     134 GAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCT 193

Db      66 btgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 111
           :     :  :  :  :  :  :  :  :  :  :     :  :  :  :
Qy     194 TGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTC 239


RESULT   14
ID   Q51787 standard; DNA; 39 BP.
AC   Q51787;
DT   20-DEC-1993  (first entry)
DE   Mixed oligonucleotide #19 encodes ballast constituent.
KW   Fusion protein; ballast constituent; pro-insulin production;
KW   recombinant protein production; HMG CoA reductase;
KW   human 3-hydroxy-3-methylglutaryl-coenzyme A-reductase;
KW   mixed oligonucleotide; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   repeat_unit     4..6
FT                   /*tag=  a
FT                   /rpt_type= tandem
FT                   /note= "can be repeated y times, where y is 4-11,
FT                   provided that y+z is 6-12"
FT   repeat_region   4..36
FT                   /*tag=  b
FT                   /note= "(DCD)11"
FT   repeat_unit     36..38
FT                   /*tag=  c
FT                   /rpt_type= tandem
FT                   /note= "can be repeated z times, where z is 1-4,
FT                   provided that y+z is 6-12; N stands for
FT                   identical or different nucleotides,
FT                   excluding stop codons"
PN   US5227293-A.
PD   13-JUL-1993.
PF   29-AUG-1989; 399874.
PR   29-AUG-1989; US-399874.
```

```
PR   23-APR-1992; US-838221.
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARH ) HOECHST AG.
PI   Habermann P,  Seed B,  Stengelin S,  Uhlmann E,  Ulmer W;
DR   WPI; 93-235119/29.
PT   Fusion proteins for prodn. of e.g. pro-insulin - comprise gene
PT   for desired protein and oligo-nucleotide(s) encoding ballast
PT   protein
PS   Claim 9; Column 30; 22pp; English.
CC   This preferred mixed oligonucleotide encodes a ballast constituent
CC   and is inserted between a regulatory region and the structural gene
CC   encoding a desired protein, esp. pro-insulin. The short ballast
CC   component improves protease resistance of the fusion protein while
CC   still allowing the desired protein to adopt its correct conformation
CC   prior to cleavage of the ballast constituent.
SQ   Sequence   39 BP;    1 A;    11 C;    1 G;    1 T;

  Query Match           10.7%;  Score 25;  DB 7;  Length 39;
  Best Local Similarity  27.3%;  Pred. No. 8.42e-02;
  Matches     9;  Conservative   20;  Mismatches   4;  Indels   0;  Gaps   0;

Db      2 tgdcddcddcddcddcddcddcddcddcddcddcdd 34
          ||:|:: :: ::|: |::|::|::|:: ::|::
Cp    158 TGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAA 126


RESULT   15
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    5 A;    10 C;    0 G;    0 T;

  Query Match           10.7%;  Score 25;  DB 12;  Length 114;
  Best Local Similarity   8.6%;  Pred. No. 8.42e-02;
  Matches     8;  Conservative   23;  Mismatches  62;  Indels   0;  Gaps   0;

Db     18 bnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnnbnnbnnbnnbnnbn 77
          :  :  :  :||   :      :  :  :  :  :  :  :  :  :  :  :
Cp    144 CCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGG 85

Db     78 bcacnnbnnbnnbnnbnnbnnbnnbnnbcacca 110
          :||   :  :  :. :  :  :  :   | |||
Cp     84 TCAAGAGGAGGACAAGGGGCAGCCCAGACCCCA 52


Search completed: Sun Oct 24 18:00:18 1999
Job time : 78 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 18:00:35 1999; MasPar time 20.77 Seconds
974.454 Million cell updates/sec
Tabular output not generated.

Title: >US-09-092-296-7
Description: (51-284) from US09092296.seq
Perfect Score: 234
N.A. Sequence: 51 ATGGGGTCTGGGCTGCCCCT.....ACCATGTTGTCTGCAACACA 284
Comp: TACCCCAGACCCGACGGGGA.....TGGTACAACAGACGTTGTGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-issued
1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 7.399; Variance 4.241; scale 1.745

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	48	20.5	7218	2	US-08-232-	Sequence 14, Applicati	3.43e-17
c	2	29	12.4	215	1	US-08-238-	Sequence 5, Applicatio	2.36e-05
c	3	28	12.0	965	3	US-08-388-	Sequence 22, Applicati	9.00e-05
	4	27	11.5	215	1	US-08-238-	Sequence 5, Applicatio	3.38e-04
c	5	23	9.8	2277	2	US-08-676-	Sequence 2, Applicatio	5.68e-02
c	6	23	9.8	2277	3	US-08-676-	Sequence 2, Applicatio	5.68e-02
	7	22	9.4	75	4	PCT-US95-1	Sequence 99, Applicati	1.95e-01
	8	22	9.4	82	4	PCT-US95-1	Sequence 97, Applicati	1.95e-01
	9	22	9.4	965	3	US-08-388-	Sequence 22, Applicati	1.95e-01
c	10	21	9.0	65	1	US-08-471-	Sequence 145, Applicat	6.54e-01
c	11	21	9.0	68	1	US-08-471-	Sequence 143, Applicat	6.54e-01
	12	21	9.0	74	4	PCT-US95-1	Sequence 94, Applicati	6.54e-01
	13	21	9.0	74	4	PCT-US95-1	Sequence 100, Applicat	6.54e-01
c	14	21	9.0	74	4	PCT-US95-1	Sequence 100, Applicat	6.54e-01
c	15	21	9.0	75	4	PCT-US95-1	Sequence 99, Applicati	6.54e-01
	16	21	9.0	81	4	PCT-US95-1	Sequence 98, Applicati	6.54e-01
	17	21	9.0	81	4	PCT-US95-1	Sequence 92, Applicati	6.54e-01
c	18	21	9.0	81	4	PCT-US95-1	Sequence 98, Applicati	6.54e-01
c	19	21	9.0	82	4	PCT-US95-1	Sequence 97, Applicati	6.54e-01
	20	21	9.0	906	3	US-08-847-	Sequence 41, Applicati	6.54e-01
	21	21	9.0	906	3	US-09-031-	Sequence 41, Applicati	6.54e-01
c	22	21	9.0	906	3	US-08-847-	Sequence 40, Applicati	6.54e-01
c	23	21	9.0	906	3	US-09-031-	Sequence 40, Applicati	6.54e-01
	24	21	9.0	908	3	US-09-031-	Sequence 39, Applicati	6.54e-01
	25	21	9.0	908	3	US-08-847-	Sequence 39, Applicati	6.54e-01
c	26	21	9.0	908	3	US-09-031-	Sequence 37, Applicati	6.54e-01
c	27	21	9.0	908	3	US-08-847-	Sequence 37, Applicati	6.54e-01
	28	21	9.0	909	3	US-09-031-	Sequence 26, Applicati	6.54e-01
	29	21	9.0	909	3	US-08-847-	Sequence 26, Applicati	6.54e-01
c	30	21	9.0	909	3	US-09-031-	Sequence 25, Applicati	6.54e-01
c	31	21	9.0	909	3	US-08-847-	Sequence 25, Applicati	6.54e-01
	32	21	9.0	911	3	US-09-031-	Sequence 24, Applicati	6.54e-01
	33	21	9.0	911	3	US-08-847-	Sequence 24, Applicati	6.54e-01
c	34	21	9.0	911	3	US-08-847-	Sequence 22, Applicati	6.54e-01
c	35	21	9.0	911	3	US-09-031-	Sequence 22, Applicati	6.54e-01
	36	21	9.0	5235	3	US-09-031-	Sequence 36, Applicati	6.54e-01
	37	21	9.0	5235	3	US-08-847-	Sequence 36, Applicati	6.54e-01
c	38	21	9.0	5235	3	US-09-031-	Sequence 35, Applicati	6.54e-01
c	39	21	9.0	5235	3	US-08-847-	Sequence 35, Applicati	6.54e-01
	40	21	9.0	5503	3	US-09-031-	Sequence 34, Applicati	6.54e-01
	41	21	9.0	5503	3	US-08-847-	Sequence 34, Applicati	6.54e-01
c	42	21	9.0	5503	3	US-08-847-	Sequence 32, Applicati	6.54e-01
c	43	21	9.0	5503	3	US-09-031-	Sequence 32, Applicati	6.54e-01
c	44	20	8.5	81	4	PCT-US95-1	Sequence 92, Applicati	2.14e+00
	45	20	8.5	2496	1	US-08-254-	Sequence 2, Applicatio	2.14e+00

ALIGNMENTS

```
RESULT    1
ID   US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC   xxxxxx
DT
DE   Sequence 14, Application US/08232463
CC   Sequence 14, Application US/08232463
CC   Patent No. 5670367
CC    GENERAL INFORMATION:
CC      APPLICANT:  DORNER, F.
CC      APPLICANT:  SCHEIFLINGER, F.
CC      APPLICANT:  FALKNER, F. G.
CC      TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
CC      NUMBER OF SEQUENCES:  52
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Foley & Lardner
CC        STREET:  1800 Diagonal Road, Suite 500
CC        CITY:  Alexandria
CC        STATE:  VA
CC        COUNTRY:  USA
CC        ZIP:  22313-0299
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/232,463
CC        FILING DATE:
CC        CLASSIFICATION:  435
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  US/07/935,313
CC        FILING DATE:
CC        APPLICATION NUMBER:  EP 91 114 300.6
CC        FILING DATE:  26-AUG-1991
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  BENT, Stephen A.
CC        REGISTRATION NUMBER:  29,768
CC        REFERENCE/DOCKET NUMBER:  30472/114 IMMU
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703)836-9300
CC        TELEFAX:  (703)683-4109
CC        TELEX:  899149
CC    INFORMATION FOR SEQ ID NO:  14:
CC      SEQUENCE CHARACTERISTICS:
```

```
CC       LENGTH:  7218 base pairs
CC       TYPE:  nucleic acid
CC       STRANDEDNESS:  single
CC       TOPOLOGY:  linear
CC     IMMEDIATE SOURCE:
CC       CLONE:  pTZgpt-Fls
SQ   SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

  Query Match           20.5%;  Score 48;  DB 2;  Length 7218;
  Best Local Similarity   1.8%;  Pred. No. 3.43e-17;
  Matches    4;  Conservative  130;  Mismatches  86;  Indels   0;  Gaps   0;

Db   1221 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1280
           :: :::::: :::::::::  :::::::  :  ::: : :    :    ::   : :
Qy     63 CTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATG 122

Db   1281 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1340
            :::: :  ::   ::       ::::::::  :   :::::::: :   ::: ::::
Qy    123 ACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTC 182

Db   1341 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1400
           ::    :: :::    :::: :::::::::: :::::::::    ::  : :: ::
Qy    183 CTGGAATTGCTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACC 242

Db   1401 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440
           :::: :: : :    ::::  : :: : :: :::| | ||
Qy    243 CTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAACA 282


RESULT    2
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC    INFORMATION FOR SEQ ID NO:  5:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  215 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  protein
CC      FEATURE:
CC        NAME/KEY:  misc_feature
CC        LOCATION:  1..215
CC        OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC        OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           12.4%;  Score 29;  DB 1;  Length 215;
  Best Local Similarity  15.1%;  Pred. No. 2.36e-05;
  Matches    26;  Conservative   69;  Mismatches  75;  Indels   2;  Gaps   2;

Db       1 MTNVTMSSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNK 60
             :| :|::: ::::: |:  ::|:::|   :::  :|   |  :   :||: :  : |: :
Cp     235 CTGGTCCCTGAAGGGAGATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTG 176

Db      61 YSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSY 120
             :  : : |  :|||:|: : :  :   ::: :|::       :  |  :   :  :::
Cp     175 GAC-TCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACC 117

Db     121 -GSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTKKHAKNSADGKVGS 171
            |::    |   :|| |   : ::  |: :: ::   ::: : :|:|::|:
Cp     116 CGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGGC 65


RESULT    3
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
CC        TELEFAX:  212-838-3884
CC    INFORMATION FOR SEQ ID NO:  22:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  965 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  unknown
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.
```

```
  Query Match           12.0%;  Score 28;  DB 3;  Length 965;
  Best Local Similarity  16.3%;  Pred. No. 9.00e-05;
  Matches   26;  Conservative   65;  Mismatches  69;  Indels   0;  Gaps   0;

Db    771 AARTGGGTKKGURHUVHVSGGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKR 830
          ||: | | ::| :  : ::| ::  :| | ::     :  ::|: :| :|| :|    ::
Cp    225 AAGGGAGATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGG 166

Db    831 GRVTMADTSSNSRSSVTAADTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHUVKDMTSSSS 890
           :: : :|:   :  :  | | :::|::   :::: |  : || :|:::: :  :|:  :
Cp    165 AGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTC 106

Db    891 ASVGDRVTTCRSSTTHGNGNTYYWYKGKAKYRVSNRSGVS 930
            |:::   :::   | |  ::: : :   ::: ::|::
Cp    105 CATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 66


RESULT    4
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC    INFORMATION FOR SEQ ID NO:  5:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  215 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  protein
CC      FEATURE:
CC        NAME/KEY:  misc_feature
CC        LOCATION:  1..215
CC        OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC        OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           11.5%;  Score 27;  DB 1;  Length 215;
  Best Local Similarity  20.0%;  Pred. No. 3.38e-04;
  Matches   22;  Conservative   42;  Mismatches  45;  Indels   1;  Gaps   1;

Db     18 CNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAK 77
          | :: : :| |  ::| :   :  |:|: :| :::  : |:  : :|   :|   :|| :
Qy     89 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG-AAGGAGT 147

Db     78 THYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
            :::|:| :: |: : :| ::  :   :: | |:   :||: :|: : |:
Qy    148 CTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAA 197


RESULT    5
ID   US-08-676-967-2 STANDARD; DNA; UNC; 2277 BP.
AC   xxxxxx
DT
DE   Sequence 2, Application US/08676967
CC   Sequence 2, Application US/08676967
CC   Patent No. 5747317
CC    GENERAL INFORMATION:
CC      APPLICANT:  COLLINS, KATHLEEN
CC      TITLE OF INVENTION:  Human Telomerase
CC      NUMBER OF SEQUENCES:  10
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Science & Technology Law Group
CC        STREET:  268 Bush Street, Suite 3200
CC        CITY:  San Francisco
CC        STATE:  CA
CC        COUNTRY:  USA
CC        ZIP:  94104
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/676,967
CC        FILING DATE:
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Osman Ph.D., Richard A
CC        REGISTRATION NUMBER:  36,627
CC        REFERENCE/DOCKET NUMBER:  UCB96-055
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415)343-4341
CC        TELEFAX:  (415)343-4342
CC    INFORMATION FOR SEQ ID NO:  2:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  2277 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  double
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  cDNA
SQ   SEQUENCE 2277 BP; 511 A; 212 C; 395 G; 216 T; 943 OTHER.

  Query Match           9.8%;  Score 23;  DB 2;  Length 2277;
  Best Local Similarity  40.0%;  Pred. No. 5.68e-02;
  Matches   18;  Conservative   13;  Mismatches  14;  Indels   0;  Gaps   0;

Db   1036 WSNGARGARGARGARYTNGGNGARYTNYTNCARCARTTYGGNGAR 1080
          :: ||:  :||:||:   |  ||::| :|  |:|||:||:   ||:
Cp    224 AGGGAGATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAA 180


RESULT    6
ID   US-08-676-974-2 STANDARD; DNA; UNC; 2277 BP.
AC   xxxxxx
DT
DE   Sequence 2, Application US/08676974
CC   Sequence 2, Application US/08676974
CC   Patent No. 5770422
CC    GENERAL INFORMATION:
CC      APPLICANT:  COLLINS, KATHLEEN
CC      TITLE OF INVENTION:  Human Telomerase
CC      NUMBER OF SEQUENCES:  10
```

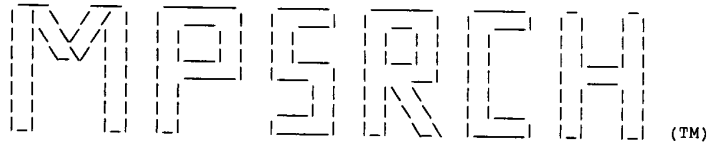
```
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Science & Technology Law Group
CC        STREET:  268 Bush Street, Suite 3200
CC        CITY:  San Francisco
CC        STATE:  CA
CC        COUNTRY:  USA
CC        ZIP:  94104
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/676,974
CC        FILING DATE:
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Osman Ph.D., Richard A
CC        REGISTRATION NUMBER:  36,627
CC        REFERENCE/DOCKET NUMBER:  UCB96-055
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415)343-4341
CC        TELEFAX:  (415)343-4342
CC    INFORMATION FOR SEQ ID NO:  2:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  2277 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  double
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  cDNA
SQ   SEQUENCE 2277 BP; 511 A; 212 C; 395 G; 216 T; 943 OTHER.

  Query Match            9.8%;  Score 23;  DB 3;  Length 2277;
  Best Local Similarity  40.0%;  Pred. No. 5.68e-02;
  Matches    18;  Conservative   13;  Mismatches  14;  Indels   0;  Gaps   0;

Db   1036 WSNGARGARGARGARYTNGGNGARYTNYTNCARCARTTYGGNGAR 1080
          :: ||:  :||:||:   |  ||::| :|  |:||:||:   ||:
Cp    224 AGGGAGATGGAGGAGGAGGCAGAGCTTTCAAGCAATTCCAGGAA 180


RESULT    7
ID   PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.
AC   xxxxxx
DT
DE   Sequence 99, Application PC/TUS9511934
CC   Sequence 99, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  99:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  75 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 75 BP; 1 A; 1 C; 7 G; 5 T; 61 OTHER.

  Query Match            9.4%;  Score 22;  DB 4;  Length 75;
  Best Local Similarity  8.7%;  Pred. No. 1.95e-01;
  Matches     6;  Conservative   19;  Mismatches  44;  Indels   0;  Gaps   0;

Db      5 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 64
            |   : : : : : : : : : : : : : : : : :     : : :
Qy     54 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 113

Db     65 NBGGTTGTG 73
           :|| | ||
Qy    114 CCGGGTATG 122


RESULT    8
ID   PCT-US95-11934-97 STANDARD; DNA; UNC; 82 BP.
AC   xxxxxx
DT
DE   Sequence 97, Application PC/TUS9511934
CC   Sequence 97, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  97:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  82 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 82 BP; 1 A; 2 C; 10 G; 8 T; 61 OTHER.

  Query Match            9.4%;  Score 22;  DB 4;  Length 82;
  Best Local Similarity  8.7%;  Pred. No. 1.95e-01;
  Matches     6;  Conservative   19;  Mismatches  44;  Indels   0;  Gaps   0;
```

```
Db      12 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 71
           |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :      :  :  :
Qy      54 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 113

Db      72 NBGGTTGTG 80
            :|| | ||
Qy     114 CCGGGTATG 122


RESULT    9
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
CC        TELEFAX:  212-838-3884
CC    INFORMATION FOR SEQ ID NO:  22:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  965 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  unknown
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

  Query Match              9.4%;  Score 22;  DB 3;  Length 965;
  Best Local Similarity   20.0%;  Pred. No. 1.95e-01;
  Matches   18;  Conservative   39;  Mismatches  32;  Indels   1;  Gaps   1;

Db     874 TVTVSSHUVKDMTSSSSASVGDRVTTCRSSTTHGNGNTYYWYKGKAKYRVSNRSGVSRSG 933
           |: :: ::::::|::::   |: :| |  : |  :    |:::::| |::  :  |: ::
Qy      52 TGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACC-CTCCTTGGCAGCTCACATGGAACA 110

Db     934 SGSGTDYTTSSDATYYCGTHARTGGTKVKG 963
           :|:  :  |  :| :: | :| ||  :::|
Qy     111 GGGCCGGGTATGACTTTGCAACTGAAGCTG 140


RESULT   10
ID   US-08-471-052A-145 STANDARD; DNA; UNC; 65 BP.
AC   xxxxxx
DT
DE   Sequence 145, Application US/08471052A
CC   Sequence 145, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  145:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  65 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 65 BP; 3 A; 3 C; 3 G; 2 T; 54 OTHER.

  Query Match              9.0%;  Score 21;  DB 1;  Length 65;
  Best Local Similarity   13.8%;  Pred. No. 6.54e-01;
  Matches    9;  Conservative   16;  Mismatches  40;  Indels   0;  Gaps   0;

Db       1 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNA 60
           || ||:  :  :  :  :     :  :  :  :  :  :  :  :     :  :  :  |
Cp     178 CTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATA 119

Db      61 CCTGG 65
           || ||
Cp     118 CCCGG 114


RESULT   11
ID   US-08-471-052A-143 STANDARD; DNA; UNC; 68 BP.
AC   xxxxxx
DT
DE   Sequence 143, Application US/08471052A
CC   Sequence 143, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
```

```
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  143:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  68 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 68 BP; 3 A; 3 C; 5 G; 3 T; 54 OTHER.

  Query Match             9.0%;  Score 21;  DB 1;  Length 68;
  Best Local Similarity  13.8%;  Pred. No. 6.54e-01;
  Matches     9;  Conservative   16;  Mismatches  40;  Indels   0;  Gaps   0;

Db       4 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNA 63
           || ||:  :  :  :  :     :  :  :  :  :  :  :  :     :  :  :  |
Cp     178 CTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATA 119

Db      64 CCTGG 68
           || ||
Cp     118 CCCGG 114


RESULT   12
ID   PCT-US95-11934-94 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 94, Application PC/TUS9511934
CC   Sequence 94, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  94:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 3 A; 4 C; 3 G; 1 T; 63 OTHER.

  Query Match             9.0%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity   4.6%;  Pred. No. 6.54e-01;
  Matches     3;  Conservative   19;  Mismatches  43;  Indels   0;  Gaps   0;

Db       5 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 64
           |  :  :  :  :  :  :  :  :  :  :  :  :  :  :     :  :  :  :  :
Qy      60 GGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGT 119

Db      65 NBNAC 69
             : ||
Qy     120 ATGAC 124


RESULT   13
ID   PCT-US95-11934-100 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 100, Application PC/TUS9511934
CC   Sequence 100, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  100:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 6 A; 6 C; 1 G; 1 T; 60 OTHER.

  Query Match             9.0%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity  13.7%;  Pred. No. 6.54e-01;
  Matches    10;  Conservative   17;  Mismatches  46;  Indels   0;  Gaps   0;
```

```
Db      1 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNV 60
          || ||:  :  :  :  :  :  :     :     :  :  :  :  :  :  :  :
Qy    192 CTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCTCCACCAT 251

Db     61 NNVNNACCACAAC 73
            :  | | ||||
Qy    252 GCAAGATCTCAAC 264


RESULT   14
ID   PCT-US95-11934-100 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 100, Application PC/TUS9511934
CC   Sequence 100, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  100:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 6 A; 6 C; 1 G; 1 T; 60 OTHER.

  Query Match             9.0%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity   7.5%;  Pred. No. 6.54e-01;
  Matches     5;  Conservative   19;  Mismatches  43;  Indels   0;  Gaps   0;

Db      5 AVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVN 64
          |:  :  :  :  :  :  :  :     :  :  :  :  :  :  :  :  :  :  :
Cp    125 AGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 66

Db     65 NACCACA 71
            | |  ||
Cp     65 CAGCCCA 59


RESULT   15
ID   PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.
AC   xxxxxx
DT
DE   Sequence 99, Application PC/TUS9511934
CC   Sequence 99, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  99:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  75 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 75 BP; 1 A; 1 C; 7 G; 5 T; 61 OTHER.

  Query Match             9.0%;  Score 21;  DB 4;  Length 75;
  Best Local Similarity   7.6%;  Pred. No. 6.54e-01;
  Matches     5;  Conservative   18;  Mismatches  43;  Indels   0;  Gaps   0;

Db      4 AGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNB 63
          ||   :  :  :  :  :  :     :  :  :  :  :  :  :  :  :  :  :
Cp    149 AGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAG 90

Db     64 NNBGGT 69
            :|||
Cp     89 GAGGGT 84


Search completed: Sun Oct 24 18:00:58 1999
Job time : 23 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:50:56 1999; MasPar time 454.93 Seconds
1205.245 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-7
Description: (51-284) from US09092296.seq
Perfect Score: 234
N.A. Sequence: 51 ATGGGGTCTGGGCTGCCCCT.....ACCATGTTGTCTGCAACACA 284
Comp: TACCCCAGACCCGACGGGGA.....TGGTACAACAGACGTTGTGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est2 20:gb_est20
21:gb_est21 22:gb_est22 23:gb_est23 24:gb_est24
25:gb_est25 26:gb_est26 27:gb_est27 28:gb_est28
29:gb_est29 30:gb_est3 31:gb_est4 32:gb_est5 33:gb_est6
34:gb_est7 35:gb_est8 36:gb_est9 37:gb_gss1 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 9.804; Variance 2.109; scale 4.649

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	73	31.2	328	23	AI136523	UI-R-C2p-nq-e-02-0-UI.	7.25e-84
	2	57	24.4	252	17	AA754459	97SN1787 Rice Immature	2.49e-56
c	3	41	17.5	252	17	AA754459	97SN1787 Rice Immature	2.28e-30
	4	33	14.1	247	17	AA754458	97SN1784 Rice Immature	2.20e-18
c	5	33	14.1	247	17	AA754458	97SN1784 Rice Immature	2.20e-18
c	6	32	13.7	2275	20	AF034173	AF034173 Human mRNA (T	5.89e-17
c	7	27	11.5	1287	20	AF038250	AF038250 Human mRNA (T	4.09e-10
	8	27	11.5	2275	20	AF034173	AF034173 Human mRNA (T	4.09e-10
c	9	26	11.1	190	21	AI010074	EST204525 Normalized r	8.11e-09
	10	25	10.7	176	19	T94049	ye33f07.r1 Stratagene	1.51e-07
c	11	25	10.7	398	8	T52782	ya79f01.r1 Stratagene	1.51e-07
c	12	25	10.7	582	37	B04965	cSRL-51A6-u cSRL flow	1.51e-07
	13	23	9.8	364	34	W33870	mc56b03.r1 Soares mous	4.22e-05
c	14	23	9.8	397	30	R52894	yh01d09.r1 Soares infa	4.22e-05
	15	23	9.8	400	28	AI554565	tn25f06.x1 NCI_CGAP_Br	4.22e-05
c	16	23	9.8	420	31	H56592	yt88b09.r1 Soares_pine	4.22e-05
	17	23	9.8	426	8	T15255	crs855 lambdaZAPST Ric	4.22e-05
c	18	23	9.8	459	13	AA475002	vh08h04.r1 Soares mous	4.22e-05
	19	23	9.8	461	22	AI086444	qf25e01.x1 NCI_CGAP_Br	4.22e-05
	20	23	9.8	547	25	AI337590	qw91a08.x1 NCI_CGAP_Br	4.22e-05
	21	23	9.8	1287	20	AF038250	AF038250 Human mRNA (T	4.22e-05
c	22	22	9.4	201	28	AI535335	UI-R-C3-sw-e-12-0-UI.s	6.28e-04
	23	22	9.4	259	38	B81136	CIT-HSP-2015F16.TFC CI	6.28e-04
	24	22	9.4	308	8	D40392	RICS2342A Rice shoot O	6.28e-04
	25	22	9.4	354	35	W84061	mf26e11.r1 Soares mous	6.28e-04
c	26	22	9.4	404	17	AA730403	nw42c07.s1 NCI_CGAP_Ew	6.28e-04
c	27	22	9.4	408	17	T46345	9608 Lambda-PRL2 Arabi	6.28e-04
	28	22	9.4	424	33	W13114	ma86e10.r1 Soares mous	6.28e-04
	29	22	9.4	433	34	W79098	zd75h10.r1 Soares_feta	6.28e-04
	30	22	9.4	441	14	C28493	C28493 Rice callus cDN	6.28e-04
	31	22	9.4	459	27	AI442028	sa66f12.y1 Gm-c1004 Gl	6.28e-04
	32	22	9.4	481	42	AQ445188	GSSTc01459 Trypanosoma	6.28e-04
	33	22	9.4	513	40	AQ235722	HS_2015_B2_C08_T7 CIT	6.28e-04
	34	22	9.4	602	23	AI166962	xylem.est.742 Poplar x	6.28e-04
c	35	22	9.4	634	22	AI063013	GH02423.5prime GH Dros	6.28e-04
	36	22	9.4	1025	37	B12587	F22O11-Sp6.1 IGF Arabi	6.28e-04
	37	21	9.0	317	19	F10052	HSC39H122 normalized i	8.54e-03
	38	21	9.0	354	26	AI398058	NCSM1A11T7 Subtracted	8.54e-03
	39	21	9.0	422	33	W07189	za94b07.r1 Soares_feta	8.54e-03
c	40	21	9.0	516	31	H41536	yp71c01.s1 Soares adul	8.54e-03
c	41	21	9.0	524	16	AA620971	af88f08.s1 Soares_test	8.54e-03
	42	21	9.0	550	20	AA898701	NCM6G7T7 Mycelial Neur	8.54e-03
c	43	21	9.0	567	30	R61539	yh16f01.s1 Soares infa	8.54e-03
	44	21	9.0	587	13	AA428548	zw47d06.r1 Soares_tota	8.54e-03
	45	21	9.0	630	39	AQ201160	RPCI11-46K18.TJ RPCI11	8.54e-03

ALIGNMENTS

RESULT 1
LOCUS AI136523 328 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C2p-nq-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone UI-R-C2p-nq-e-02-0-UI 3', mRNA sequence.
ACCESSION AI136523
NID g3637300
VERSION AI136523.1 GI:3637300
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 328)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 14, 1998 this sequence version replaced gi:1877567.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.
FEATURES Location/Qualifiers

```
    source          1. .328
                    /organism="Rattus norvegicus"
                    /strain="Sprague-Dawley"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p
                    library is a subtracted library derived from the UI-R-C1
                    library, which is a subtracted library derived from the
                    UI-R-C0 library.  The UI-R-C0 library consisted of a
                    mixture of individually tagged normalized  libraries
                    constructed from rat placenta, adult lung, brain, liver,
                    kidney,  heart, spleen, ovary, muscle, 8, 12 and 18-day
                    embryo. The tag is a string of  3-5 nucleotides present
                    between the Not I site and the oligo-dT track which allows
                    identification of the library of origin of a clone within
                    the mixture. The subtracted library (UI-R-C2p) was
                    constructed as follows: PCR amplified cDNA inserts from
                    UI-R-C1 clones from which 3' ESTs had been derived was
                    used as a driver in a hybridization with the UI-R-C1
                    library in the form of single-stranded circles. The
                    remaining single-stranded circles (subtracted library) was
                    purified by hydroxyapatite column chromatography,
                    converted to double-stranded circles and electroporated
                    into DH10B bacteria (Life Technologies) to generate the
                    UI-R-C2p library. This procedure has been previously
                    described (Bonaldo, Lennon and Soares, Genome Research 6:
                    791-806, 1996)"
                    /db_xref="taxon:10116"
                    /clone="UI-R-C2p-nq-e-02-0-UI"
                    /clone_lib="UI-R-C2p"
                    /dev_stage="adult"
                    /lab_host="DH10B (Life Technologies)"
BASE COUNT       62 a     77 c     98 g     91 t
ORIGIN

  Query Match           31.2%;  Score 73;  DB 23;  Length 328;
  Best Local Similarity  73.0%;  Pred. No. 7.25e-84;
  Matches   116;  Conservative    0;  Mismatches  43;  Indels   0;  Gaps   0;

Db    142 TGCAGGTAAGGTGGTGTGGTGGCCCTTTATGATGAAGGGTGACATTGGTCCCCGGTGAGA 201
          |||||  ||  |||||| | |  | |  ||| || ||||||||  |||||||  |  | ||
Cp    279 TGCAGACAACATGGTGTTGAGATCTTGCATGGTGGAGGGTGACGCTGGTCCCTGAAGGGA 220

Db    202 GGTGGAGGAGGAGGCAGATCTTTTGGAGCATGTCCAGAAAGCCGGAGTCTTGGGAGGTCT 261
          | ||||||||||||||||| |||||  ||||  ||||| |||| ||| || | |||||  |
Cp    219 GATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGGAGGAAT 160

Db    262 TGGCCTGAGATGCCTCGGTCAGTTTTACTTGCAAAGTCA 300
          | | | || | | |||  |||| || | ||||||||||
Cp    159 TTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCA 121


RESULT    2
LOCUS       AA754459      252 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1787, mRNA sequence.
ACCESSION   AA754459
NID         g2801165
VERSION     AA754459.1  GI:2801165
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797457.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES            Location/Qualifiers
     source         1. .252
                    /organism="Oryza sativa"
                    /cultivar="Milyang23"
                    /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                    XhoI; Directional cDNA library inserted into lambda ZAPII
                    vector  at 5'end with EcoRI and 3' end with Xho I site."
                    /db_xref="taxon:4530"
                    /map="6"
                    /clone="97SN1787"
                    /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                    /tissue_type="Immature Seed"
                    /dev_stage="5 days after pollination"
                    /lab_host="E. coli SOLR"
BASE COUNT       5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match           24.4%;  Score 57;  DB 17;  Length 252;
  Best Local Similarity  10.9%;  Pred. No. 2.49e-56;
  Matches    23;  Conservative  108;  Mismatches  78;  Indels   2;  Gaps   2;

Db     22 SYBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWB 81
          ::: :  ::  |  ::  : :::   ::| :|:|  : :  :  | :  : : |: ::::
Qy     69 CTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTG 128

Db     82 VBNTKVDVGNHTRCSRWRBVTRMAHYHDYTNCBBYNNNDYHMWHBBMY-BBTGCMTCTMW 140
          :   ::::| : : ::::::|   :: ::   :::    ::  ::::: ::| |:|  ::
Qy    129 CAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAA 188

Db    141 CWBHYNTKCTASGWHTSTNYDVKSSTNTWGVTBSYDKSMHGYWCSBBVKYHTKVSTTRAT 200
           ::::     |: ::|:  : ::::|  : :|::::::: | :|: :::::  ::|  |
Qy    189 TTGCTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCTCCAC 248

Db    201 RSYTCVRK-YCVMWMTKKVVKKYHVVBBGCH 230
           : |::: :| :::   ::::::: ::|||:
Qy    249 CATGCAAGATCTCAACACCATGTTGTCTGCA 279


RESULT    3
LOCUS       AA754459      252 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1787, mRNA sequence.
ACCESSION   AA754459
NID         g2801165
VERSION     AA754459.1  GI:2801165
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797457.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
```

```
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .252
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /map="6"
                      /clone="97SN1787"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                      /tissue_type="Immature Seed"
                      /dev_stage="5 days after pollination"
                      /lab_host="E. coli SOLR"
BASE COUNT        5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match           17.5%;  Score 41;  DB 17;  Length 252;
  Best Local Similarity  10.3%;  Pred. No. 2.28e-30;
  Matches    16;  Conservative   79;  Mismatches  58;  Indels   2;  Gaps   2;

Db      27 GNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWBVBNTK 86
           | :::: : ::|   :::   :  |:|: : : : |: :| : ::  :|: :::::  :
Cp     221 GAGATGGAGGAGGAGGCAGAGCTTTTCAAGCAAT-TCCAG-GAAGCTGGACTCATAGGAG 164

Db      87 VDVGNHTRCSRWRBVTRMAHYHDYTNCBBYNNNDYHMWHBBMYBBTGCMTCTMWCWBHYN 146
           :::  :  | ::: : :: :::::  |:::  : :::::::  :   |  | :: ::::
Cp     163 GAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCA 104

Db     147 TKCTASGWHTSTNYDVKSSTNTWGVTBSYDKSMHG 181
           |:  |: :  :   :::::|  :|: :  : :::|
Cp     103 TGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAG 69


RESULT    4
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
VERSION     AA754458.1  GI:2801164
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797455.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .247
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /map="6"
                      /clone="97SN1784"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                      /tissue_type="Immature Seed"
                      /dev_stage="5 days after pollination"
                      /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match           14.1%;  Score 33;  DB 17;  Length 247;
  Best Local Similarity  15.4%;  Pred. No. 2.20e-18;
  Matches    20;  Conservative   55;  Mismatches  55;  Indels   0;  Gaps   0;

Db       9 VWRGCCCCBAWMNKHTHMMTBBWCCVRRVGTTTNNGKHNGRTTTWNDCSDNAHCRYTVBW 68
           :::| || : :: :: :   : :| :::: |    |     |||  : :: |:  :|:::
Qy     109 CAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCT 168

Db      69 YYARSKYGYGTBYYSWNVDTNTGGTGVGKTTVNVHSGWNNRCSNSVVYVWBTAYCDYBHY 128
           : ::::  | ::::: ::  || | :    :  : |    |  : :::::| :|::::
Qy     169 ATGAGTCCAGCTTCCTGGAATTGCTTGAAAAGCTCTGCCTCCTCCTCCATCTCCCTTCAG 228

Db     129 BDRANHVDDT 138
           :::  :: :|
Qy     229 GGACCAGCGT 238


RESULT    5
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
VERSION     AA754458.1  GI:2801164
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797455.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .247
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /map="6"
                      /clone="97SN1784"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
```

```
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match           14.1%;  Score 33;  DB 17;  Length 247;
  Best Local Similarity  14.7%;  Pred. No. 2.20e-18;
  Matches    24;  Conservative   71;  Mismatches  66;  Indels   2;  Gaps   1;

Db      34 VRRVGTTTNNGKHNGRTTTWNDCSDNAHCRYTVBWYYARSKYGYGTBYYSWNVDTNTGGT 93
           ::::| |   |:: |:   : : ::  :|:  : :::   :   | :     ::|  |
Cp     223 GGGAGATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGGAG 164

Db      94 GVGKTTVNVHSGWNNRCSNSVVYVWBTAYCDYBHYBDRANHVDDTRCTNDRGYCNYTASD 153
           |: :||: ::: :  : : :: ::::  :|: :: :::| ::::  |  :  :  :| ::
Cp     163 GAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCA 104

Db     154 NGTSATKRVTGYDKTDSDCGGGCWRKVTYGSSBYBRCGVNVMV 196
           ||:|  ::||:   ::: || |::::   :: : : |: :::
Cp     103 TGTGA--GCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGGCAG 63


RESULT    6
LOCUS       AF034173     2275 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
VERSION     AF034173.1  GI:2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045115.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match           13.7%;  Score 32;  DB 20;  Length 2275;
  Best Local Similarity  11.2%;  Pred. No. 5.89e-17;
  Matches    13;  Conservative   66;  Mismatches  35;  Indels   2;  Gaps   2;

Db    1480 RYKRWKRRKGRRKRMTGMYKRMYRAMMMAMCAMMACWWYYWKMRGMKKCWKYRKYKKYTS 1539
           ::::::::|::: :   ::  : |: :|:  ::|  :   : :|: :|:: :: ::
Cp     218 ATGGAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCT-GGACTCATAGGAGGAAT 160

Db    1540 TYYKSWSRWYWYTTYTYWYCWCCT-SMKSASCAMMRWMGYMGSRSSRSYWGYWGSM 1594
           |: :::::: :   : :::|: || :::  :||:::::: :: : :: :::|:: ::
Cp     159 TTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCA 104


RESULT    7
LOCUS       AF038250     1287 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF038250 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon9, mRNA sequence.
ACCESSION   AF038250
NID         g2815880
VERSION     AF038250.1  GI:2815880
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1287)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045085.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .1287
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon9"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      349 a    219 c    293 g    361 t     65 others
ORIGIN

  Query Match           11.5%;  Score 27;  DB 20;  Length 1287;
  Best Local Similarity  45.6%;  Pred. No. 4.09e-10;
  Matches    26;  Conservative   15;  Mismatches  16;  Indels   0;  Gaps   0;

Db     432 GKVGGVBGASTCMTHYNCCDCKTCGSAGVTVNHHDWSMAGGAARAGCTTCTCTCGCA 488
           |::||::||: |   : ||:    |:||:|: : ::: ||| |:|||| ||  |||
Cp     246 GGAGGGTGACGCTGGTCCCTGAAGGGAGATGGAGGAGGAGGCAGAGCTTTTCAAGCA 190


RESULT    8
LOCUS       AF034173     2275 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
VERSION     AF034173.1  GI:2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045115.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
```

```
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match           11.5%;  Score 27;  DB 20;  Length 2275;
  Best Local Similarity  17.8%;  Pred. No. 4.09e-10;
  Matches    19;  Conservative   53;  Mismatches  33;  Indels   2;  Gaps   2;

Db   1534 YKKYTSTYYKSWSRWYWYTTYTYWYCWCCTSMKSASCAMMRWMGYMGSRSSRSYWGYWGS 1593
            :::: :|:::::: ::: :| :|:  |: ||::   : |:::  |::|::   :   ::|:
Qy     70 TTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGC 129

Db   1594 MSGCYGMTKRYYRYSWTGWTKWTTWWYMWSMTRWTMTTYTTWWTWRT 1640
            :  |:|: : : : :  |: ::|| : : :: ::|: |: |:: ::|
Qy    130 AA-CTGAAGCTGAAGGAGTCTTTTCTGA-CAAATTCCTCCTATGAGT 174


RESULT    9
LOCUS       AI010074      190 bp    mRNA            EST       15-JUN-1998
DEFINITION  EST204525 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
            RLUBT38 3' end, mRNA sequence.
ACCESSION   AI010074
NID         g3223906
VERSION     AI010074.1  GI:3223906
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 190)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151091.

            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .190
                     /organism="Rattus sp."
                     /note="Organ: lung; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
                     /db_xref="taxon:10118"
                     /map="21q"
                     /clone="RLUBT38"
                     /clone_lib="Normalized rat lung, Bento Soares"
BASE COUNT       39 a     51 c     59 g     41 t
ORIGIN

  Query Match           11.1%;  Score 26;  DB 21;  Length 190;
  Best Local Similarity  69.7%;  Pred. No. 8.11e-09;
  Matches    46;  Conservative    0;  Mismatches  20;  Indels   0;  Gaps   0;

Db    125 TGCAGGTAAGGTGGTGTGGTGGCCCTTTATGATGAAGGGTGACATTGGTCCCCGGTGAGA 184
          |||||  ||  |||||| | |  | |  ||| || |||||||| |||||||| |  | ||
Cp    279 TGCAGACAACATGGTGTTGAGATCTTGCATGGTGGAGGGTGACGCTGGTCCCTGAAGGGA 220

Db    185 GGTGGA 190
          | ||||
Cp    219 GATGGA 214


RESULT   10
LOCUS       T94049        176 bp    mRNA            EST       24-MAR-1995
DEFINITION  ye33f07.r1 Stratagene lung (#937210) Homo sapiens cDNA clone
            IMAGE:119557 5', mRNA sequence.
ACCESSION   T94049
NID         g727537
VERSION     T94049.1  GI:727537
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 196
            High quality sequence stops: 155 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 196   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 155.
FEATURES             Location/Qualifiers
     source          1. .176
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. normal lung.  Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="GDB:487846"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:119557"
                     /clone_lib="Stratagene lung (#937210)"
                     /sex="male"
                     /dev_stage="72 years"
                     /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT       48 a     46 c     39 g     39 t      4 others
ORIGIN

  Query Match           10.7%;  Score 25;  DB 19;  Length 176;
  Best Local Similarity 100.0%;  Pred. No. 1.51e-07;
  Matches    25;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 TCAACACCATGTTGTCTGCAACACA 25
          |||||||||||||||||||||||||
Qy    260 TCAACACCATGTTGTCTGCAACACA 284


RESULT   11
LOCUS       T52782        398 bp    mRNA            EST       06-FEB-1995
DEFINITION  ya79f01.r1 Stratagene ovary (#937217) Homo sapiens cDNA clone
            IMAGE:67897 5', mRNA sequence.
ACCESSION   T52782
NID         g654642
VERSION     T52782.1  GI:654642
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
```

```
            Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Other_ESTs: ya79f01.s1
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1135
            High qality sequence stops: 352 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1135   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 352.
FEATURES              Location/Qualifiers
     source           1. .398
                      /organism="Homo sapiens"
                      /note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
                      Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                      Total ovary tissue, normal, caucasian. Average insert
                      size: 0.8 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
                      GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3'"
                      /db_xref="GDB:504154"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:67897"
                      /clone_lib="Stratagene ovary (#937217)"
                      /sex="female"
                      /dev_stage="49 year old"
                      /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT       95 a    125 c     94 g     82 t      2 others
ORIGIN

  Query Match           10.7%;  Score 25;  DB 8;  Length 398;
  Best Local Similarity  90.3%;  Pred. No. 1.51e-07;
  Matches    28;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db    158 GGAGATGGCGGAGGAGTCAGAGCTTTTGAAG 188
          ||||||||| ||||||| ||||||||||| |||
Cp    222 GGAGATGGAGGAGGAGGCAGAGCTTTTCAAG 192


RESULT   12
LOCUS       B04965        582 bp    DNA             GSS       26-JUN-1996
DEFINITION  cSRL-51A6-u cSRL flow sorted Chromosome 11 specific cosmid Homo
            sapiens genomic clone cSRL-51A6, genomic survey sequence.
ACCESSION   B04965
NID         g1414243
VERSION     B04965.1  GI:1414243
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
            Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S. .,
            Harris,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M.,
            Kupfer,K. and Garner,H.R.
  TITLE     Genomic Sequence Sampled Map of Chromosome 11
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Evans GA, Shane Probst
            McDermott Center for Human Growth and Development
            University of Texas Southwestern Medical Center At Dallas
            5323 Harry Hines Blvd, Dallas TX 75235-8591
            Tel: 214-648-1600
            Fax: 214-648-1666
            Email: gevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
            PCR PRimers
            FORWARD: AGACGAAACACATAGACGAG
            BACKWARD: TTCAAACTCAAAAAGCTCTG
            Seq primer: T7
            Class: cosmid ends
            High quality sequence stop: 582.
FEATURES              Location/Qualifiers
     source           1. .582
                      /organism="Homo sapiens"
                      /note="Vector: sCos-1; Human Chromosome 11 specific cosmid
                      library prepared from flow sorted human  Chromosome 11
                      derived from Chinese Hampster Ovary (CHO) monochromosomal
                      somatic cell hybrid, J1"
                      /db_xref="taxon:9606"
                      /clone="cSRL-51A6"
                      /clone_lib="cSRL flow sorted Chromosome 11 specific
                      cosmid"
                      /sex="female"
                      /cell_type="chimeric hamster somatic cell hybrid"
BASE COUNT      145 a    130 c    135 g    138 t     34 others
ORIGIN

  Query Match           10.7%;  Score 25;  DB 37;  Length 582;
  Best Local Similarity  74.2%;  Pred. No. 1.51e-07;
  Matches    46;  Conservative    0;  Mismatches  15;  Indels   1;  Gaps   1;

Db    232 CTGGCATGGTGGCAGGTGACTTTAATCCCAGCACTTGGGATGCAGGGGCAGGCAGAGCTT 291
          || ||||||||| ||||||  |  |||| | |   | |||| ||| | ||||||||||||
Cp    256 CTTGCATGGTGGAGGGTGACGCTGGTCCCTGAAGG-GAGATGGAGGAGGAGGCAGAGCTT 198

Db    292 TT 293
          ||
Cp    197 TT 196


RESULT   13
LOCUS       W33870        364 bp    mRNA            EST       13-MAY-1996
DEFINITION  mc56b03.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:352493 5', mRNA sequence.
ACCESSION   W33870
NID         g1315794
VERSION     W33870.1  GI:1315794
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 18, 1995 this sequence version replaced gi:811194.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:224293
            Seq primer: ETPrimer
            High quality sequence stop: 360.
FEATURES              Location/Qualifiers
     source           1. .364
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "
/db_xref="taxon:10090"
/clone="IMAGE:352493"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"

BASE COUNT 100 a 73 c 87 g 104 t
ORIGIN

Query Match 9.8%; Score 23; DB 34; Length 364;
Best Local Similarity 79.5%; Pred. No. 4.22e-05;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 282 AACTGAAAGGGAAGGAGGCTTTTCCCACATATTTCTCCT 320
||||||| ||||||| |||||| ||| ||| |||||
Qy 130 AACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCT 168

RESULT 14
LOCUS R52894 397 bp mRNA EST 18-MAY-1995
DEFINITION yh01d09.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:41873 5', mRNA sequence.
ACCESSION R52894
NID g814796
VERSION R52894.1 GI:814796
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 397)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Sep 21, 1992 this sequence version replaced gi:275991.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 299
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: M13RP1
High quality sequence stop: 299.

FEATURES Location/Qualifiers
source 1. .397
/organism="Homo sapiens"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not I; Site_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="GDB:414414"
/db_xref="taxon:9606"
/map="9"
/clone="IMAGE:41873"
/clone_lib="Soares infant brain 1NIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"

BASE COUNT 115 a 74 c 100 g 104 t 4 others
ORIGIN

Query Match 9.8%; Score 23; DB 30; Length 397;
Best Local Similarity 78.0%; Pred. No. 4.22e-05;
Matches 32; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 7 GAGGAGAAGTCAGAACATTACCAGTATTTCCAGGAAGGTGG 47
|||||| || |||| | || | || | |||||||||| |||
Cp 215 GAGGAGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTGG 175

RESULT 15
LOCUS AI554565 400 bp mRNA EST 23-MAR-1999
DEFINITION tn25f06.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2168675 3' similar to TR:Q16664 Q16664 PROTEIN A-1. [1] ;, mRNA sequence.
ACCESSION AI554565
NID g4486928
VERSION AI554565.1 GI:4486928
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 400)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT On May 7, 1998 this sequence version replaced gi:3121680.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.

FEATURES Location/Qualifiers
source 1. .400
/organism="Homo sapiens"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/map="19p12-p13.1; p12"
/clone="IMAGE:2168675"
/clone_lib="NCI_CGAP_Brn25"

/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
BASE COUNT 77 a 117 c 124 g 82 t
ORIGIN

Query Match 9.8%; Score 23; DB 28; Length 400;
Best Local Similarity 74.5%; Pred. No. 4.22e-05;
Matches 35; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Db    268 AAGCCCTTGCACTTTCTTCACCCCCCGTCAGGGAGCAGCGTCTCCCT 314
          |||| ||  | | | || || | ||| ||||||| ||||||| ||||
Qy    198 AAGCTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCACCCT 244
```

Search completed: Sun Oct 24 17:58:39 1999
Job time : 463 secs.

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:19:23 1999; MasPar time 370.71 Seconds
1345.905 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-3
Description: (1-180) from US09092296.seq
Perfect Score: 180
N.A. Sequence: 1 CAGGAGCGCAGTGGCCACTA.....CTTGAAAAGCTCTGCCTCCT 180
Comp: GTCCTCGCGTCACCGGTGAT.....GAACTTTTCGAGACGGAGGA

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl58
1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi
Database: genbank111
17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 9.304; Variance 4.612; scale 2.018

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	90	50.0	47323	31	AC005937	Homo sapiens clone UWG	8.42e-49
	2	32	17.8	216021	31	HUAC004787	Homo sapiens Chromosom	7.20e-06
	3	29	16.1	965	25	AR024229	Sequence 22 from paten	5.53e-04
	4	28	15.6	7218	25	I66494	Sequence 14 from paten	2.26e-03
	5	27	15.0	215	25	I28278	Sequence 5 from patent	9.06e-03
c	6	27	15.0	215	25	I28278	Sequence 5 from patent	9.06e-03
c	7	26	14.4	965	25	AR024229	Sequence 22 from paten	3.54e-02
	8	26	14.4	74371	31	AC005369	Homo sapiens chromosom	3.54e-02
c	9	26	14.4	216021	31	HUAC004787	Homo sapiens Chromosom	3.54e-02
	10	25	13.9	565	25	E04076	gDNA encoding envelope	1.35e-01
	11	25	13.9	60966	31	AC003030	Homo sapiens chromosom	1.35e-01
c	12	24	13.3	60	25	A62989	Sequence 1 from Patent	5.03e-01
c	13	24	13.3	1056	23	MVU87256	Mustela vison GT dinuc	5.03e-01
	14	23	12.8	1056	23	MVU87256	Mustela vison GT dinuc	1.82e+00
	15	23	12.8	3290	29	HUMPFUCAS	H.sapiens fucosidase p	1.82e+00
c	16	23	12.8	175902	20	AC006992	Homo sapiens clone NH0	1.82e+00
c	17	22	12.2	30	25	A62994	Sequence 6 from Patent	6.36e+00
c	18	22	12.2	108	21	D87227	Trypanosoma cruzi mRNA	6.36e+00
c	19	22	12.2	1738	29	HUMGOLGINB	Human (clone SY2/10) g	6.36e+00
c	20	22	12.2	1945	24	CHKCONN45	Chicken connexin-45 mR	6.36e+00
c	21	22	12.2	2180	21	DROSYT	D.melanogaster synapto	6.36e+00
	22	22	12.2	2273	32	MMBAKEXN	Mus musculus Bak gene,	6.36e+00
c	23	22	12.2	3088	11	E10775	DNA encoding part of A	6.36e+00
c	24	22	12.2	3088	25	AR014574	Sequence 1 from patent	6.36e+00
c	25	22	12.2	6640	29	D63997	Homo sapiens mRNA for	6.36e+00
	26	22	12.2	67919	20	AC006542	WORKING DRAFT SEQUENCE	6.36e+00
	27	22	12.2	82098	31	AC006252	Homo sapiens 3p21.1 co	6.36e+00
c	28	22	12.2	124700	22	AC005558	Drosophila melanogaste	6.36e+00
c	29	22	12.2	133457	30	AC003999	Human PAC clone DJ1139	6.36e+00
	30	22	12.2	135039	31	AC006060	Homo sapiens 3p22-8 PA	6.36e+00
	31	22	12.2	151187	19	HS460J8	Human DNA sequence ***	6.36e+00
c	32	22	12.2	175339	31	AC005772	Homo sapiens chromosom	6.36e+00
	33	22	12.2	203418	19	AC004947	Homo sapiens clone DJ1	6.36e+00
c	34	22	12.2	241407	20	AC003059	Mouse Chromosome 10 BA	6.36e+00
c	35	21	11.7	65	25	I41365	Sequence 145 from pate	2.16e+01
	36	21	11.7	631	30	HSATP2A1S2	Human Ca2+ ATPase of f	2.16e+01
c	37	21	11.7	1000	30	HSTAL1DNA	H.sapiens tal-1 DNA.	2.16e+01
	38	21	11.7	1755	29	HSZ97017	Homo sapiens mRNA for	2.16e+01
	39	21	11.7	3425	29	AB006623	Homo sapiens mRNA for	2.16e+01
	40	21	11.7	46213	28	SPBC18H10	S.pombe chromosome II	2.16e+01
c	41	21	11.7	56804	30	HS77N19	Human DNA sequence fro	2.16e+01
c	42	21	11.7	74371	31	AC005369	Homo sapiens chromosom	2.16e+01
c	43	21	11.7	118995	31	AC005368	Homo sapiens chromosom	2.16e+01
c	44	21	11.7	127027	30	HS461P17	Human DNA sequence fro	2.16e+01
	45	21	11.7	187966	31	AC006487	Homo sapiens chromosom	2.16e+01

ALIGNMENTS

RESULT 1
LOCUS AC005937 47323 bp DNA PRI 05-NOV-1998
DEFINITION Homo sapiens clone UWGC:370M23.002 from 6p21, complete sequence.
ACCESSION AC005937
NID g3845393
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 47323)
AUTHORS Janer,M., Guillaudeux,T., Vu,Q., Kutyavin,T., Harter,H. and Geraghty,D.E.
TITLE Large scale sequence analysis of the human MHC class I region
JOURNAL Unpublished (1998)
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
REFERENCE 2 (bases 1 to 47323)
AUTHORS Geraghty,D.E. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhcrc.org)
COMMENT Overlapping Sequences:
5': UWGC:370m23.013 (Genbank Accession: AC005530)
3': UWGC:y67c112 (Genbank Accession: AC004211)

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than

1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 75.5%
DS or two chemistry coverage: 98.9%
Single stranded regions: 3

Sequence Validation:
This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BglII		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
1069.11	1050.00	889.55	866.00	30541.40	30653.00
20320.67	20855.00	1050.18	1015.00	3279.08	3231.00
2171.50	2147.00	7268.78	7196.00		
2560.20	2531.00	10085.80	9992.00		
4335.42	4269.00	11212.78	11131.00		
2698.62	2628.00				
1927.50	1887.00				
3130.46	3090.00				
2166.69	2129.00				
2044.67	2005.00				

```
FEATURES             Location/Qualifiers
     source          1. .47323
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21"
                     /sub_clone="UWGC:370M23.002"
                     /clone_lib="Research Genetics BAC Library"
     repeat_region   3647. .3932
                     /rpt_family="Alu"
     repeat_region   complement(4999. .5277)
                     /rpt_family="Alu"
     repeat_region   6285. .6572
                     /rpt_family="Alu"
     repeat_region   complement(6972. .7050)
                     /rpt_family="MLT1"
     repeat_region   7286. .7584
                     /rpt_family="Alu"
     repeat_region   complement(8164. .8609)
                     /rpt_family="Alu"
     repeat_region   complement(21287. .21895)
                     /rpt_family="Alu"
     repeat_region   22715. .22957
                     /rpt_family="Alu"
     repeat_region   25510. .25802
                     /rpt_family="Alu"
     repeat_region   27835. .28010
                     /rpt_family="MER20"
     repeat_region   31295. .31594
                     /rpt_family="Alu"
     repeat_region   33515. .33767
                     /rpt_family="Alu"
     repeat_region   34223. .34290
                     /rpt_family="MIR"
     repeat_region   37372. .37648
                     /rpt_family="Alu"
     repeat_region   38526. .38700
                     /rpt_family="MER3"
     repeat_region   39583. .40010
                     /rpt_family="Alu"
     repeat_region   40046. .40156
                     /rpt_family="Alu"
     repeat_region   43194. .43372
                     /rpt_family="MER5"
     variation       43325
                     /note="clonal variation with 3' overlapping clone"
     variation       44149
                     /note="clonal variation with 3' overlapping clone"
     variation       44451
                     /note="clonal variation with 3' overlapping clone"
     variation       44537
                     /note="clonal variation with 3' overlapping clone"
     variation       44814
                     /note="clonal variation with 3' overlapping clone"
     variation       44965
                     /note="clonal variation with 3' overlapping clone"
     variation       45760
                     /note="clonal variation with 3' overlapping clone"
     variation       45900
                     /note="clonal variation with 3' overlapping clone"
     variation       46851
                     /note="clonal variation with 3' overlapping clone"
     variation       46859
                     /note="clonal variation with 3' overlapping clone"
     variation       47032
                     /note="clonal variation with 3' overlapping clone"
     variation       47240. .47256
                     /note="clonal variation with 3' overlapping clone -
                     insertion of 17bp repeat"
BASE COUNT    11556 a  11489 c  12284 g  11994 t
ORIGIN

  Query Match           50.0%;  Score 90;  DB 31;  Length 47323;
  Best Local Similarity  98.9%;  Pred. No. 8.42e-49;
  Matches    91;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db  35465 CAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCT 35524
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Qy     78 CAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAGTTCCTCCT 137

Db  35525 ATGAGTCCAGCTTCCTGGAATTGCTTGAAAAG 35556
          ||||||||||||||||||||||||||||||
Qy    138 ATGAGTCCAGCTTCCTGGAATTGCTTGAAAAG 169


RESULT    2
LOCUS       HUAC004787 216021 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
NID         g3337381
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES                Location/Qualifiers
     source             1. .216021
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /map="#16q21-22"
                        /clone="A-952F10"
     STS                27765. .27872
                        /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                        sapiens"
                        /db_xref="dbSTS:G02280"
     STS                73826. .73943
                        /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                        sapiens"
                        /db_xref="dbSTS:G02122"
     STS                175801. .175945
                        /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                        /db_xref="dbSTS:G09703"
     STS                175810. .175945
                        /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                        /db_xref="dbSTS:G09935"
     STS                199463. .199572
                        /note="9824, WI-3555, Chr. 16, Homo sapiens"
                        /db_xref="dbSTS:G04338"
BASE COUNT    60960 a  51778 c  49172 g  53987 t    124 others
ORIGIN

  Query Match           17.8%;  Score 32;  DB 31;  Length 216021;
  Best Local Similarity  11.6%;  Pred. No. 7.20e-06;
  Matches    15;  Conservative   68;  Mismatches  45;  Indels   1;  Gaps   1;

Db    1454 RGKKKYMWKSMRRARRRGSAGKKKKKYYYYYCYYYYYYYYCMGRAMMAAAWYYKRRSCMA 1513
           :|::::::  : : ::    |: :: ::::: ::::::::  |: ::   :: : ::|
Qy      10 AGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTC 69

Db    1514 WYYYMRGRARKTYYYMARRRGCARSYKKYMAMYWAAWTTWCAAAAMAAMYTTYYWMMWYY 1573
           :: : : : ::  ::  : :  | :::: :|::: |:  : || :  ::||::: :
Qy      70 ACATGGAACAGGGCCGGGTATGA-CTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAA 128

Db    1574 KKWMYTCCT 1582
           :::::||||
Qy     129 GTTCCTCCT 137


RESULT    3
LOCUS       AR024229      965 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 22 from patent US 5795961.
ACCESSION   AR024229
NID         g3977523
VERSION     AR024229.1  GI:3977523
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and
            Kitamura,K.
  TITLE     Recombinant human anti-Lewis b antibodies
  JOURNAL   Patent: US 5795961-A 22 18-AUG-1998;
FEATURES                Location/Qualifiers
     source             1. .965
                        /organism="unknown"
BASE COUNT      192 a    170 c    226 g    205 t    172 others
ORIGIN

  Query Match           16.1%;  Score 29;  DB 25;  Length 965;
  Best Local Similarity  20.6%;  Pred. No. 5.53e-04;
  Matches    22;  Conservative   47;  Mismatches  37;  Indels   1;  Gaps   1;

Db     857 VRGRSYDSDGGDYWGGTTVTVSSHTVKDMTSSSSASVGDRVTTCRSSTTHGNGNTYYWYK 916
           ::| ::::::|| ::   |: :: :|::::|::::   |: :| | : | :    |:::::
Qy       4 GAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACC-CTCCTTG 62

Db     917 GKAKYRVSNRSGVSRSGSGSGTDYTTSSDATYYCGTHARTGGTKVKG 963
           | |:: :  : |: :: :|:  :  |  :| :: | :| ||  :::|
Qy      63 GCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG 109


RESULT    4
LOCUS       I66494        7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES                Location/Qualifiers
     source             1. .7218
                        /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match           15.6%;  Score 28;  DB 25;  Length 7218;
  Best Local Similarity   2.5%;  Pred. No. 2.26e-03;
  Matches     4;  Conservative   90;  Mismatches  66;  Indels   0;  Gaps   0;

Db    1061 TGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1120
           || | |::   :: :::::: :::::::::  ::::::::  :  ::: : :    :
Qy      21 TGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAG 80

Db    1121 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1180
             ::   : :  :::: :  ::    ::       ::::::::  :  ::::::::  :
Qy      81 GGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAGTTCCTCCTATG 140

Db    1181 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1220
             :::  ::::::    :: :::      :::: ::::::
Qy     141 AGTCCAGCTTCCTGGAATTGCTTGAAAAGCTCTGCCTCCT 180


RESULT    5
LOCUS       I28278        215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
VERSION     I28278.1  GI:1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
```

```
          Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M,, Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="unknown"
BASE COUNT       15 a      8 c     25 g     26 t    141 others
ORIGIN

  Query Match           15.0%;  Score 27;  DB 25;  Length 215;
  Best Local Similarity  20.0%;  Pred. No. 9.06e-03;
  Matches    22;  Conservative   42;  Mismatches  45;  Indels   1;  Gaps   1;

Db      18 CNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAK 77
           | :: : :| |  ::| :   :  |:|: :| :::  : |:  : :|   :|   :|| :
Qy      58 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG-AAGGAGT 116

Db      78 THYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
             :::|:| :: |: : :| ::  :   :: | |:   :||: :|: : |:
Qy     117 CTTTTCTGACAAGTTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAA 166


RESULT    6
LOCUS       I28278        215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
VERSION     I28278.1  GI:1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="unknown"
BASE COUNT       15 a      8 c     25 g     26 t    141 others
ORIGIN

  Query Match           15.0%;  Score 27;  DB 25;  Length 215;
  Best Local Similarity  13.8%;  Pred. No. 9.06e-03;
  Matches    22;  Conservative   66;  Mismatches  69;  Indels   2;  Gaps   2;

Db       1 MTNVTMSSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNK 60
           :| : :::  :: : |:  ::|:::|   : : |    : :|:|:  ||:::  :: : :
Cp     158 ATTCCAGGAAGCTGGACTCATAGGAGGAACTTGTCAGAAAAGACTCCTTCAGCTTC-AGT 100

Db      61 YSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTD-GNRSGADS 119
           :::|   :    :   : ::: :| :   ::  :|::: |:| :::     : | ::||
Cp      99 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 40

Db     120 YGSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTK 158
            |::: |  :  :       |:::: :   ::|::  |:
Cp      39 AGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTG 1


RESULT    7
LOCUS       AR024229      965 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 22 from patent US 5795961.
ACCESSION   AR024229
NID         g3977523
VERSION     AR024229.1  GI:3977523
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and
            Kitamura,K.
  TITLE     Recombinant human anti-Lewis b antibodies
  JOURNAL   Patent: US 5795961-A 22 18-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .965
                     /organism="unknown"
BASE COUNT      192 a    170 c    226 g    205 t    172 others
ORIGIN

  Query Match           14.4%;  Score 26;  DB 25;  Length 965;
  Best Local Similarity  20.0%;  Pred. No. 3.54e-02;
  Matches    18;  Conservative   41;  Mismatches  30;  Indels   1;  Gaps   1;

Db     781 GTRHTVHVSGGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKRGRVTMADTSS 840
           || :|::::||: :|:| | ::: |:   ::  :  |: || :    |::|:: :|: ::
Cp      93 GTCATACCCGGCCCTGT-TCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 35

Db     841 NSRSSVTAADTAVYYCVRGRSYDSDGGDYW 870
             ::::  |   : : : |::::::| :::
Cp      34 CAGCCCAGACCCCATAGTGGCCACTGCGCT 5


RESULT    8
LOCUS       AC005369    74371 bp    DNA             PRI       01-AUG-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 119j3 (LBNL H175), complete
            sequence.
ACCESSION   AC005369
NID         g3367505
VERSION     AC005369.1  GI:3367505
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 74371)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 74371)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 74371)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .74371
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="119j3"
                     /chromosome="5"
                     /note="LBNL H175"
     repeat_region   893. .1030
                     /rpt_family="Alu"
     repeat_region   2295. .2438
                     /rpt_family="Alu"
     repeat_region   2818. .2859
                     /note="(GT)21"
                     /rpt_type=tandem
                     /rpt_unit=GT
```

misc_feature join(3246. .3410,3721. .3828)
/standard_name="RLF"
/note="65% & 69% protein identity GenPept:U22377"
repeat_region 3431. .3724
/rpt_family="Alu"
repeat_region 3707. .3728
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A
repeat_region 4366. .4661
/rpt_family="Alu"
repeat_region 5327. .5602
/rpt_family="Alu"
repeat_region 6586. .6956
/rpt_family="L1"
repeat_region 6647. .6684
/note="(CA)19"
/rpt_type=tandem
/rpt_unit=CA
repeat_region 7113. .7373
/rpt_family="Alu"
repeat_region complement(7830. .8185)
/standard_name="possible repeat"
repeat_region 8258. .8503
/rpt_family="Alu"
repeat_region 9070. .9387
/rpt_family="Alu"
repeat_region complement(9740. .9845)
/rpt_family="MER42"
repeat_region complement(10440. .11015)
/rpt_family="Alu"
repeat_region 11950. .12250
/rpt_family="Alu"
repeat_region 12057. .12085
/note="(A)29"
/rpt_type=tandem
/rpt_unit=A
repeat_region 12365. .12645
/rpt_family="Alu"
repeat_region 13727. .13750
/note="(AC)12"
/rpt_type=tandem
/rpt_unit=AC
repeat_region 13783. .14024
/rpt_family="L1"
repeat_region 14175. .14470
/rpt_family="Alu"
repeat_region complement(14906. .15259)
/standard_name="possible repeat"
repeat_region 15300. .15613
/rpt_family="Alu"
repeat_region 16671. .16690
/note="(T)20"
/rpt_type=tandem
/rpt_unit=T
repeat_region complement(16675. .16977)
/rpt_family="Alu"
repeat_region complement(16993. .17085)
/rpt_family="MER42"
repeat_region complement(17678. .18276)
/rpt_family="Alu"
repeat_region 19305. .19583
/rpt_family="Alu"
repeat_region 19914. .19945
/note="(GTTT)8"
/rpt_type=tandem
/rpt_unit=GTTT
repeat_region complement(19943. .20222)
/rpt_family="Alu"
repeat_region 21202. .21496
/rpt_family="Alu"
repeat_region 21736. .22035
/rpt_family="Alu"
repeat_region 22017. .22038
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A
repeat_region 22321. .22457
/rpt_family="Alu"
repeat_region complement(22638. .22981)
/standard_name="possible repeat"
repeat_region 23022. .23326
/rpt_family="Alu"
repeat_region 23473. .23761
/rpt_family="Alu"
repeat_region 23744. .23767
/note="(A)24"
/rpt_type=tandem
/rpt_unit=A
repeat_region complement(24639. .24694)
/rpt_family="MER42"
repeat_region complement(25349. .25713)
/rpt_family="Alu"
repeat_region complement(25727. .26471)
/rpt_family="Alu"
repeat_region 27191. .27477
/rpt_family="Alu"
repeat_region 27774. .28057
/rpt_family="Alu"
repeat_region 28040. .28066
/note="(A)27"
/rpt_type=tandem
/rpt_unit=A
repeat_region 28712. .28930
/rpt_family="MER20"
misc_feature complement(28769. .28838)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region 28987. .29214
/rpt_family="Alu"
misc_feature join(29485. .29559,29598. .29779)
/note="95% identity dbSTS:G14522 (SHGC-11312)"
misc_feature 29495. .29976
/standard_name="AI027942"
/note="100% identity EST ov84a10.x1"
misc_feature complement(30401. .30536)
/note="GRAIL 2 excellent exon, frame 2"
repeat_region complement(30682. .30733)
/rpt_family="MIR"
misc_feature complement(31573. .31724)
/note="GRAIL 2 excellent exon, frame 1"
misc_feature complement(32159. .32232)
/note="GRAIL 2 excellent exon, frame 2"
repeat_region 32388. .32488
/rpt_family="MLT1"
repeat_region 32617. .32908
/rpt_family="Alu"
repeat_region 32977. .33088
/rpt_family="MLT1"
repeat_region complement(33670. .33785)
/rpt_family="Alu"
misc_feature complement(34021. .34144)
/note="GRAIL 2 excellent exon, frame 2"
misc_feature complement(35238. .35331)
/note="GRAIL 2 excellent exon, frame 1"
repeat_region 36392. .36663
/rpt_family="Alu"
misc_feature 36901. .37222
/note="100% identity EST ou55c09.x1"
/db_xref="dbEST:AI025011"
STS 36901. .37164
/standard_name="STSG-9983"
/db_xref="dbSTS:G26554"
misc_feature complement(37595. .37654)
/note="GRAIL 2 excellent exon, frame 0"
misc_feature complement(join(38063. .38218,38462. .38578,38741. .38995,
39071. .39205,39532. .39630,39935. .40048,40300. .40410,

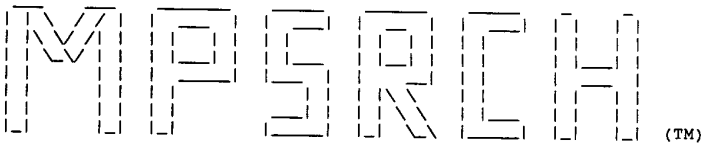
```
                    40503. .40661,41868. .41972,42103. .42225,42492. .42569,
                    44379. .44507))
                    /standard_name="histidyl-tRNA synthetase"
                    /note="78%-100% protein identity GenPept:U18937"
     misc_feature   complement(38069. .38215)
                    /note="GRAIL 2 excellent exon, frame 0"
...
Note: remainder of annotations omitted.

  Query Match           14.4%;  Score 26;  DB 31;  Length 74371;
  Best Local Similarity  17.9%;  Pred. No. 3.54e-02;
  Matches    10;  Conservative   31;  Mismatches  15;  Indels   0;  Gaps   0;

Db  15956 KSSKCWCWSYRMRMKCYSCSYCYCSSGKKYWCRCSMYWYTCYYSYKYYWSMSYCTC 16011
          :: :| |:::: : |: : : : |:::: |:::: ||:::::::::::::|||
Qy      3 GGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTC 58


RESULT    9
LOCUS       HUAC004787 216021 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
NID         g3337381
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES            Location/Qualifiers
     source         1. .216021
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="16"
                    /map="#16q21-22"
                    /clone="A-952F10"
     STS            27765. .27872
                    /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                    sapiens"
                    /db_xref="dbSTS:G02280"
     STS            73826. .73943
                    /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                    sapiens"
                    /db_xref="dbSTS:G02122"
     STS            175801. .175945
                    /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                    /db_xref="dbSTS:G09703"
     STS            175810. .175945
                    /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                    /db_xref="dbSTS:G09935"
     STS            199463. .199572
                    /note="9824, WI-3555, Chr. 16, Homo sapiens"
                    /db_xref="dbSTS:G04338"
BASE COUNT    60960 a  51778 c  49172 g  53987 t    124 others
ORIGIN

  Query Match           14.4%;  Score 26;  DB 31;  Length 216021;
  Best Local Similarity  13.1%;  Pred. No. 3.54e-02;
  Matches    13;  Conservative   52;  Mismatches  33;  Indels   1;  Gaps   1;

Db   1446 YYCSRARRRGKKKYMWKSMRRARRRGSAGKKKKKYYYYYCYYYYYYYYCMGRAMMAAAWY 1505
           ::|:: :::| :: : ::: : :::| | :::: :::::::|:|  :| :
Cp    157 TTCCAGGAAGCTGGACTCATAGGAGGAACTTGTCAGAAAAGACTCCTTCAGCTTCAGTTG 98

Db   1506 YKRRSCMAWYYYMRGRARKTYYYMARRRGCARSYKKYMA 1544
           : ::: :|: ::::|   : ::::| : | |:::: ::|
Cp     97 CAAAGTCATACCCGGCCCTGTTCCATGTG-AGCTGCCAA 60


RESULT   10
LOCUS       E04076        565 bp    RNA             PAT       29-SEP-1997
DEFINITION  gDNA encoding envelope region of type C hepatitis virus.
ACCESSION   E04076
NID         g2172286
VERSION     E04076.1  GI:2172286
KEYWORDS    JP 1992349885-A/1.
SOURCE      Hepatitis C virus.
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepatitis C-like viruses.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Morinaga,T., Chayama,K., Kumada,H. and Ichikawa,Y.
  TITLE     NUCLEIC ACID FRAGMENT WITHIN ENVELOPE REGION OF HEPATITIS C VIRUS
            AND METHOD FOR DETECTING THE SAME
  JOURNAL   Patent: JP 1992349885-A 1 04-DEC-1992;
            TEIJIN LTD
COMMENT     OS   Hepatitis C virus
            PN   JP 1992349885-A/1
            PD   04-DEC-1992
            PF   29-MAY-1991 JP 1991152169
            PI   MORINAGA TSUTAE, CHAYAMA KAZUAKI, KUMADA HIROMITSU, PI
            ICHIKAWA YATARO
            PC   C12N15/10,C12Q1/68,C12Q1/70//C12N15/11;
            CC   strandedness: Single;
            CC   topology: Linear;.
FEATURES            Location/Qualifiers
     source         1. .565
                    /organism="Hepatitis C virus"
                    /db_xref="taxon:11103"
BASE COUNT       60 a     93 c    107 g     85 t    220 others
ORIGIN

  Query Match           13.9%;  Score 25;  DB 25;  Length 565;
  Best Local Similarity  34.1%;  Pred. No. 1.35e-01;
  Matches    30;  Conservative   29;  Mismatches  28;  Indels   1;  Gaps   1;

Db    400 GGRTCYGTYTTYCTYRTYTCYCABYTGTTTACCYTCTMBMCYMGSMDGYAYGDGACRDYR 459
          ||:||:|   | | : |:  :| ::| || |||:||  :  ::|:   :|:|: ||:: :
Qy     23 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 82

Db    460 CARGRYTGYAAYTGYTCDMTYTAYSCYG 487
          | :|::  : |:| : |:: : | :|:|
Qy     83 CCGGGTA-TGACTTTGCAACTGAAGCTG 109
```

```
RESULT   11
LOCUS       AC003030    60966 bp    DNA             PRI       06-JAN-1999
DEFINITION  Homo sapiens chromosome 19, overlapping cosmids R29828 and F25496,
            complete sequence.
ACCESSION   AC003030
NID         g4092821
VERSION     AC003030.1  GI:4092821
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
            Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S.,
            Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 2 Mb contig in 19p12 between UBA52 and
            D19S455
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1997) Human Genome Center, Joint Genome
            Institute/ Lawrence Livermore National Laboratory, 7000 East Ave.,
            Livermore, CA 94551, USA
REFERENCE   3  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE   4  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     On Jan 4, 1999 this sequence version replaced gi:2570224.
            Map and sequence oriented from p telomere to centromere. This
            accession is comprised of overlapping cosmids R29828 (bases 1 to
            40,974) and F25496 (bases 23,336 to 60,966). R29828 is separated
            from cosmid F23858 (AC004475) to the left by a sequence gap of
            approximately 14 kb, which is to be filled by sequencing a
            restriction fragment from cosmid R27236 (currently in progress).
            Cosmid F25496 is separated from cosmid R31863 to the right by an
            expected sequence gap of at least 10 kb. Additional chr 19 map and
            sequence information are available at:
            http://www-bio.llnl.gov/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1. .60966
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="R29828-F25496"
                     /chromosome="19"
                     /map="19p12 between UBA52 and D19S455."
                     /cell_line="UV5HL9-5B for F25496, and 5HL2-B for R29828"
                     /clone_lib="LL19NC02 (for F25496) and LL19NC03 (for
                     R29828) chromosome 19-specific cosmid libraries"
                     /note="Cosmid libraries constructed at LLNL from
                     flow-sorted chromosomes from human-hamster hybrid cell
                     lines UV5HL9-5B and 5HL2-B, each of which carries
                     chromosome 19 as its only human chromosome."
     repeat_region   18. .219
                     /rpt_family="AluJo"
     repeat_region   530. .824
                     /rpt_family="AluY"
     repeat_region   930. .1066
                     /rpt_family="FLAM_C"
     misc_feature    complement(1148. .1315)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: good, score: 51.000"
     repeat_region   complement(1477. .1706)
                     /rpt_family="AluJo"
     repeat_region   2105. .2406
                     /rpt_family="AluSp"
     repeat_region   2513. .2809
                     /rpt_family="AluJo"
     misc_feature    5028. .5045
                     /note="DDS similarity to overlapping ESTs:~AA963316
                     UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                     clone UI-R-E1- gi-c-06-0-UI 3'; (220. .237); 100%
                     identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                     Soares Rattus sp. cDNA clone RKIBE38 3' end; (240. .257);
                     100% identity.~"
     CDS             join(<5028. .5045,5432. .5497,6480. .6575,8155. .8249,
                     10156. .10183,10562. .10717,11845. .11991,12086. .12176,
                     13147. .13250,14614. .14679,18199. .18248,18465. .18542,
                     18635. .18746,23665. .23755,24574. .24709,37283. .37493)
                     /note="Hypothetical human protein (partial)"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="R29828_1"
                     /protein_id="AAD03161.1"
                     /db_xref="PID:g4106983"
                     /db_xref="GI:4106983"
                     /translation="WLISQQIPQFEDVKFEAASLLSELYCQENSVDAAKPLLRKAIQI
                     SQQTPYWHCRLLFQLAQLHTLEKDLVSACDLLGVGAEYARVVGSEYTRALFLLSKGML
                     LLMERKLQEVHPLLTLCGQIVENWQGNPIQKESLRVFFLVLQVTHYLDAGQVKSVKPC
                     LKQLQQCIQTISTLHDDEILPSNPADLFHWLPKEHMCVLVYLVTVMHSMQAGYLEKAQ
                     KYTDKALMQLEKLKMLDCSPILSSFQVILLEHIIMCRLVTGHKATALQEGLYCVSVNC
                     MDNAEAQFTTALRLYSLLERINPDHSFPVSSHCLRAAAFYVRGLFSFFQGRYNEAKRF
                     LRETLKMSNAEDLNRLTACSLVLLGHIFYVLGNHRESNNMVVPAMQLASKIPDMSVQL
                     WSSALLRDLNKACGNAMDAHEAAQMHQNFSQQLLQDHIEACSLPEHNLITVRVGHWEG
                     IAATRLPVAFMEGQSTTTLYSADIWKHIQEDTEGSESAWAPVFRCEIKSLRPQTRAII
                     LLPPLD"
     misc_feature    complement(5186. .5264)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 86.000"
     misc_feature    5432. .5497
                     /note="DDS similarity to overlapping ESts:~AA963316
                     UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                     clone UI-R-E1- gi-c-06-0-UI 3'; (238. .303); 88%
                     identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                     Soares Rattus sp. cDNA clone RKIBE38 3' end; (258. .323);
                     88% identity."
     repeat_region   complement(5602. .5883)
                     /rpt_family="AluJo"
     misc_feature    6480. .6575
                     /note="DDS similarity to overlapping ESTs:~AA963316
                     UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                     clone UI-R-E1- gi-c-06-0-UI 3'; (304. .399); 89%
                     identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                     Soares Rattus sp. cDNA clone RKIBE38 3' end; (324. .419);
                     89% identity.~~(6480. .6660) predicted exon, program:
                     grail2exons_human_1.3, frame: 2, quality: excellent,
                     score: 76.000"
     misc_feature    8155. .8249
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 94.000~~DDS
                     similarity to overlapping ESTs:~(8155. .8237) AA963316
                     UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                     clone UI-R-E1- gi-c-06-0-UI 3'; (400. .482); 88%
                     identity.~~(8155. .8240) AA893275 EST197078 Normalized rat
                     kidney, Bento Soares Rattus sp. cDNA clone RKIBE38 3' end;
                     (420. .505); 85% identity."
     repeat_region   8351. .8477
                     /rpt_family="AluJb"
     repeat_region   8480. .8768
                     /rpt_family="AluSc"
     misc_feature    8778. .8837
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: excellent, score: 83.000"
     repeat_region   complement(8872. .8996)
```

```
                    /rpt_family="AluJb"
     repeat_region  9052. .9196
                    /rpt_family="AluSx"
     repeat_region  9197. .9493
                    /rpt_family="AluY"
     repeat_region  9494. .9658
                    /rpt_family="AluSx"
     repeat_region  9668. .9967
                    /rpt_family="AluY"
     misc_feature   10156. .10183
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 1, quality: excellent, score: 86.000"
     misc_feature   10562. .10721
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 1, quality: excellent, score: 100.000"
     misc_feature   11845. .11991
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 0, quality: excellent, score: 97.000"
     misc_feature   12086. .12176
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 1, quality: excellent, score: 96.000"
     misc_feature   13147. .13250
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 2, quality: excellent, score: 100.000"
     repeat_region  13440. .13741
                    /rpt_family="AluSx"
     misc_feature   14614. .14679
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 0, quality: excellent, score: 100.000"
     misc_feature   16463. .16747
                    /note="DDS similarity to AA212842 mw82d12.r1 Soares mouse
                    NML Mus musculus cDNA clone 677207 5'; (200. .468); 80%
                    identity."
     repeat_region  17395. .17687
                    /rpt_family="AluJb"
     misc_feature   18199. .18248
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 0, quality: excellent, score: 89.000"
     misc_feature   18465. .18542
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 0, quality: excellent, score: 100.000"
     misc_feature   18635. .18746
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 2, quality: excellent, score: 97.000"
     repeat_region  complement(19286. .19309)
                    /rpt_family="POLY_A"
     repeat_region  19323. .19623
                    /rpt_family="AluSx"
     repeat_region  19632. .19928
                    /rpt_family="AluSx"
     repeat_region  19929. .20226
                    /rpt_family="AluY"
     repeat_region  20234. .20341
                    /rpt_family="AluJo/FLAM"
     repeat_region  complement(20344. .20655)
...
Note: remainder of annotations omitted.

  Query Match           13.9%;  Score 25;  DB 31;  Length 60966;
  Best Local Similarity  85.7%;  Pred. No. 1.35e-01;
  Matches    30;  Conservative    0;  Mismatches   5;  Indels   0;  Gaps   0;

Db  13976 TGTCCACGATGGGATCTGGGCTGCCCTTTGCCCTC 14010
          || |||| ||||| |||||||||||| ||| ||||
Qy     12 TGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTC 46


RESULT   12
LOCUS       A62989         60 bp    DNA             PAT       12-MAR-1998
DEFINITION  Sequence 1 from Patent WO9720068.
ACCESSION   A62989
NID         g3716861
VERSION     A62989.1  GI:3716861
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 60)
  AUTHORS   Oerum,H. and Seeger,C.
  TITLE     METHOD FOR GENERATING MULTIPLE DOUBLE STRANDED NUCLEIC ACIDS
  JOURNAL   Patent: WO 9720068-A 1 05-JUN-1997;
            BOEHRINGER MANNHEIM GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        0 a     30 c      0 g      0 t     30 others
ORIGIN

  Query Match           13.3%;  Score 24;  DB 25;  Length 60;
  Best Local Similarity   6.2%;  Pred. No. 5.03e-01;
  Matches     2;  Conservative   26;  Mismatches   4;  Indels   0;  Gaps   0;

Db     29 CCRRRRRRRRRRRRRRRRRRRRRRRRRRRR 60
          ||:::::::::  :::::::::: :::::: ::
Cp     63 CCAAGGAGGGTCAAGAGGAGGACAAGGGGCAG 32


RESULT   13
LOCUS       MVU87256     1056 bp    DNA             MAM       02-JAN-1999
DEFINITION  Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION   U87256
NID         g4099442
VERSION     U87256.1  GI:4099442
KEYWORDS    .
SOURCE      American mink.
  ORGANISM  Mustela vison
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
            Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /chromosome="1"
                     /map="1q"
                     /note="primers: 1167F: agcccctgcatatctacttctt, 1167R:
                     gaggatcttaccgctgttgag"
     primer_bind     98. .119
                     /standard_name="1167F"
     primer_bind     complement(300. .320)
                     /standard_name="1167R"
BASE COUNT      211 a    221 c    210 g    225 t    189 others
ORIGIN

  Query Match           13.3%;  Score 24;  DB 23;  Length 1056;
  Best Local Similarity  10.8%;  Pred. No. 5.03e-01;
  Matches     7;  Conservative   40;  Mismatches  17;  Indels   1;  Gaps   1;

Db    611 KRMRRCSRYMMMRMCVGSGTWARCCCCDKKSGSGDKSHCKSKRKYKWMDRYHBC-KSMCA 669
          ::::: : :: : :|:|:  :   || :::: : ::::  :::: ::::: ::  :::||
Cp     99 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 40

Db    670 MVRGK 674
          :::|:
Cp     39 AGGGG 35


RESULT   14
LOCUS       MVU87256     1056 bp    DNA             MAM       02-JAN-1999
DEFINITION  Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION   U87256
```

```
NID         g4099442
VERSION     U87256.1  GI:4099442
KEYWORDS    .
SOURCE      American mink.
  ORGANISM  Mustela vison
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
            Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /chromosome="1"
                     /map="1q"
                     /note="primers: 1167F: agcccctgcatatctacttctt, 1167R:
                     gaggatcttaccgctgttgag"
     primer_bind     98. .119
                     /standard_name="1167F"
     primer_bind     complement(300. .320)
                     /standard_name="1167R"
BASE COUNT      211 a    221 c    210 g    225 t    189 others
ORIGIN

  Query Match           12.8%;  Score 23;  DB 23;  Length 1056;
  Best Local Similarity  14.8%;  Pred. No. 1.82e+00;
  Matches    16;  Conservative   52;  Mismatches  39;  Indels   1;  Gaps   1;

Db    644 GDKSHCKSKRKYKWMDRYHBCKSMCAMVRGKWSAGTGCMARGWRTWDSSCTSTKGARRRK 703
          |::::|:::: :: :  :::  : |  :  :::  | |  :| :  :: |   : | :::
Qy     22 GGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 81

Db    704 GMYARRKMYGTGAAWASATAYSTGAGTTBWRVTSWKMBKYKKTGMMHH 751
          |:: :::::|    : :|  ::  || |::::   :::::: ||::::
Qy     82 GCCGGGTATGACTTTGCAACTGA-AGCTGAAGGAGTCTTTTCTGACAA 128


RESULT   15
LOCUS       HUMPFUCAS    3290 bp    DNA             PRI       08-JAN-1995
DEFINITION  H.sapiens fucosidase pseudogene.
ACCESSION   M80816
NID         g189864
VERSION     M80816.1  GI:189864
KEYWORDS    alpha-L-fucosidase; lysosomal hydrolase; lysosome; pseudogene.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3290)
  AUTHORS   Seo,H.C., Willems,P.J., Kretz,K.A., Martin,B.M. and O'Brien,J.S.
  TITLE     Fucosidosis: four new mutations and a new polymorphism
  JOURNAL   Hum. Mol. Genet. 2 (4), 423-429 (1993)
  MEDLINE   93278392
FEATURES             Location/Qualifiers
     source          1. .3290
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2q31-q32"
     gene            735. .2761
                     /gene="FUCA1P"
     misc_feature    735. .2761
                     /gene="FUCA1P"
                     /note="homology to human a-L-fucosidase cDNA; G00-119-238;
                     putative"
BASE COUNT      930 a    660 c    681 g   1019 t
ORIGIN      chromosome 2.

  Query Match           12.8%;  Score 23;  DB 29;  Length 3290;
  Best Local Similarity  79.5%;  Pred. No. 1.82e+00;
  Matches    31;  Conservative    0;  Mismatches   8;  Indels   0;  Gaps   0;

Db   1596 TGCAACATGGCAATGTCTGACTTTGCAAGTGAATCTGAA 1634
          || |||| |||   || ||||||||||| |||| |||||
Qy     73 TGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTGAA 111


Search completed: Sun Oct 24 17:25:43 1999
Job time : 380 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:32:37 1999; MasPar time 60.50 Seconds
637.452 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-3
Description: (1-180) from US09092296.seq
Perfect Score: 180
N.A. Sequence: 1 CAGGAGCGCAGTGGCCACTA.....CTTGAAAAGCTCTGCCTCCT 180
Comp: GTCCTCGCGTCACCGGTGAT.....GAACTTTTCGAGACGGAGGA

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 7.535; Variance 4.714; scale 1.598

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	178	98.9	439	60	V84366	Human stomach carcino	4.41e-99
c	2	37	20.6	91	9	Q51746	Oligonucleotide probe	1.15e-08
c	3	37	20.6	204	1	N81164	Base substituted E.co	1.15e-08
	4	34	18.9	91	9	Q51746	Oligonucleotide probe	5.55e-07
	5	33	18.3	204	1	N81164	Base substituted E.co	1.99e-06
c	6	31	17.2	114	12	Q70467	Generic DNA sequence	2.48e-05
c	7	30	16.7	114	12	Q70470	Generic DNA sequence	8.61e-05
	8	29	16.1	114	12	Q70470	Generic DNA sequence	2.96e-04
	9	29	16.1	114	12	Q70467	Generic DNA sequence	2.96e-04
	10	28	15.6	114	12	Q70469	Generic DNA sequence	1.00e-03
	11	28	15.6	114	12	Q70465	Generic DNA sequence	1.00e-03
c	12	28	15.6	114	12	Q70468	Generic DNA sequence	1.00e-03
	13	28	15.6	178	32	T76405	Human endothelin-1 an	1.00e-03
	14	27	15.0	114	12	Q70468	Generic DNA sequence	3.36e-03
c	15	27	15.0	114	12	Q70465	Generic DNA sequence	3.36e-03
	16	26	14.4	114	12	Q70472	Generic DNA sequence	1.11e-02
	17	26	14.4	114	12	Q70466	Generic DNA sequence	1.11e-02
c	18	25	13.9	39	7	Q51787	Mixed oligonucleotide	3.60e-02
c	19	25	13.9	114	12	Q70469	Generic DNA sequence	3.60e-02
c	20	24	13.3	114	12	Q70466	Generic DNA sequence	1.15e-01
c	21	24	13.3	114	12	Q70472	Generic DNA sequence	1.15e-01
	22	23	12.8	114	12	Q70473	Generic DNA sequence	3.62e-01
c	23	23	12.8	114	12	Q70471	Generic DNA sequence	3.62e-01
	24	23	12.8	172	32	T76363	Human interleukin 8 a	3.62e-01
	25	23	12.8	565	6	Q35072	HCV envelope region n	3.62e-01
c	26	22	12.2	36	2	Q11195	Ballast Constituent c	1.11e+00
	27	22	12.2	75	21	T13612	DC43 TSAR library gen	1.11e+00
	28	22	12.2	82	21	T13610	DC43 TSAR library gen	1.11e+00
	29	22	12.2	89	32	T76219	Human IL5 antisense o	1.11e+00
	30	22	12.2	114	12	Q70471	Generic DNA sequence	1.11e+00
c	31	22	12.2	114	12	Q70473	Generic DNA sequence	1.11e+00
	32	22	12.2	250	32	T76438	Substance P antisense	1.11e+00
	33	22	12.2	264	32	T76445	Substance P receptor	1.11e+00
c	34	22	12.2	501	3	N50025	Sequence encoding new	1.11e+00
c	35	22	12.2	501	3	N50026	Sequence encoding new	1.11e+00
c	36	22	12.2	3088	16	T05628	ADP ribosylation fact	1.11e+00
	37	21	11.7	66	21	T13585	TSAR-9 library genera	3.36e+00
	38	21	11.7	69	21	T13583	TSAR-9 library genera	3.36e+00
c	39	21	11.7	82	21	T13610	DC43 TSAR library gen	3.36e+00
	40	21	11.7	908	51	V63025	B. malayi ankyrin cDN	3.36e+00
c	41	21	11.7	908	51	V63024	B. malayi ankyrin nBm	3.36e+00
	42	21	11.7	909	51	V63015	D. immitis ankyrin cD	3.36e+00
c	43	21	11.7	909	51	V63014	D. immitis ankyrin nD	3.36e+00
	44	21	11.7	5235	51	V63023	D. immitis ankyrin cD	3.36e+00
	45	21	11.7	5503	51	V63021	D. immitis ankyrin cD	3.36e+00

ALIGNMENTS

RESULT 1
ID V84366 standard; cDNA to mRNA; 439 BP.
AC V84366;
DT 30-MAR-1999 (first entry)
DE Human stomach carcinoma cDNA clone HP10408.
KW Transmembrane protein; HP10408; human; stomach cancer; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 75..311
FT /*tag= a
FT /note= "cDNA comprising the coding region (minus
FT the stop codon) is claimed (Claim 3)"
PN WO9855508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI; 99-045730/04.
DR P-PSDB; W88498.
PT New human proteins containing transmembrane domains and their
PT encoding sequences - useful in the preparation of antibodies and
PT large-scale protein production, gene diagnosis, and gene therapy
PS Claim 4; Page 135; 178pp; English.
CC This is the nucleotide sequence of cDNA clone HP10408, which
CC includes a coding region (also claimed) for a novel human
CC transmembrane protein (see W88498). The clone was isolated from a
CC stomach cancer cDNA library using a signal sequence detection
CC method, and by protein synthesis by in vitro translation. The
CC encoded protein has a putative signal sequence and a putative
CC internal transmembrane domain. The invention provides nucleotide
CC sequences (see V84359-76) coding for 18 transmembrane proteins

```
CC   (see W88491-508), vectors containing such polynucleotides, and
CC   eukaryotic cells containing the vectors.  The proteins can be
CC   used as antigens or as compositions in the preparation of
CC   antibodies against the proteins.  The polynucleotides can be used
CC   as probes for gene diagnosis, and as gene sources for gene therapy
CC   and large-scale production of proteins encoded by the cDNA.  The
CC   host cells are used for the detection of ligands corresponding to
CC   the expressed proteins, and the screening of low mol.wt. medicines.
SQ   Sequence   439 BP;    89 A;    137 C;    109 G;    104 T;

  Query Match           98.9%;  Score 178;  DB 60;  Length 439;
  Best Local Similarity  99.4%;  Pred. No. 4.41e-99;
  Matches   179;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db     56 caggagcgcagtggccactatggggtctgggctgccccttgtcctcctcttgaccctcct 115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 CAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCT 60

Db    116 tggcagctcacatggaacagggccgggtatgactttgcaactgaagctgaaggagtcttt 175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 TGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTT 120

Db    176 tctgacaaattcctcctatgagtccagcttcctggaattgcttgaaaagctctgcctcct 235
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 TCTGACAAGTTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAAAAGCTCTGCCTCCT 180


RESULT    2
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match           20.6%;  Score 37;  DB 9;  Length 91;
  Best Local Similarity   1.9%;  Pred. No. 1.15e-08;
  Matches    1;  Conservative   46;  Mismatches   4;  Indels   1;  Gaps   1;

Db     11 ssvhsyyvvhvvshhhsvhhvvhhvhvsv-vvvhhvvhvvhhvhyhvyvsvc 61
          :::::::: ::: ::::::::::::::  :::::::::: ::::::::::|
Cp    133 GGAACTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGC 82


RESULT    3
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;    21 A;    47 C;    17 G;    11 T;    108 Others;

  Query Match           20.6%;  Score 37;  DB 1;  Length 204;
  Best Local Similarity   7.9%;  Pred. No. 1.15e-08;
  Matches    8;  Conservative   54;  Mismatches   39;  Indels   0;  Gaps   0;

Db     86 ymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynv 145
          :::|   : ::: : :: :  : : | ::| : :::::   :  |:::::::: :::  :
Cp    141 TCATAGGAGGAACTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGC 82

Db    146 hnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
          : :    |  :  ::|: ::    :   :::::|::::::|
Cp     81 CCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGAC 41


RESULT    4
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match           18.9%;  Score 34;  DB 9;  Length 91;
  Best Local Similarity  10.7%;  Pred. No. 5.55e-07;
  Matches    6;  Conservative   39;  Mismatches   11;  Indels   0;  Gaps   0;

Db     12 svhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctcaagc 67
          ::::::::: :::: : ::::: ::::::::::::  :: ::::  : :|| ||||
Qy     52 GACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGC 107
```

```
RESULT    5
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match           18.3%;  Score 33;  DB 1;  Length 204;
  Best Local Similarity   5.4%;  Pred. No. 1.99e-06;
  Matches     5;  Conservative    51;  Mismatches  37;  Indels    0;  Gaps    0;

Db      94 yrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynvhnhnncnc 153
           ::: :: ::::: :: |  ::    : : : :|    |:::::::::::: :: :  |
Qy       1 CAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCT 60

Db     154 ccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
             :   : : :::  :: :  : :: : ::::|
Qy      61 TGGCAGCTCACATGGAACAGGGCCGGGTATGAC 93


RESULT    6
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           17.2%;  Score 31;  DB 12;  Length 114;
  Best Local Similarity   4.6%;  Pred. No. 2.48e-05;
  Matches     5;  Conservative    30;  Mismatches  74;  Indels    0;  Gaps    0;

Db       6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnnbnn 65
             :  :  :  :  :  :     :  :  :  :  :  :  :  :||   :        :
Cp     113 CCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGG 54

Db      66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnb 114
           :     :  :  :  :  :  :  :     :  :  :  :  :|||  :
Cp      53 TCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCT 5


RESULT    7
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
```

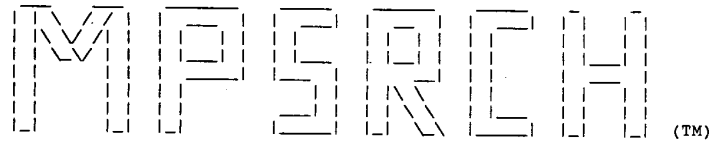
```
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    5 A;    10 C;    0 G;    0 T;

  Query Match           16.7%;  Score 30;  DB 12;  Length 114;
  Best Local Similarity   6.7%;  Pred. No. 8.61e-05;
  Matches     7;  Conservative   28;  Mismatches  69;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbcacnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :||   :  :  :  :| |  :  :  :     :  :  :  :
Cp    107 GCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGA 48

Db     63 bnnbnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbc 106
          :  :  :  :  :  :| |     :  :  :  :  :  :  :|
Cp     47 GGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTC 4


RESULT    8
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    5 A;    10 C;    0 G;    0 T;

  Query Match           16.1%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   7.5%;  Pred. No. 2.96e-04;
  Matches     7;  Conservative   25;  Mismatches  61;  Indels   0;  Gaps   0;

Db     13 cacnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnnbnnbnnbnnb 72
          ||  :  :  :  :|||  :  :  :  :  :  :  :  :        :  :  :  :
Qy      1 CAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCT 60

Db     73 nnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnb 105
          :  :| |  :  :  :  :  :  :  :
Qy     61 TGGCAGCTCACATGGAACAGGGCCGGGTATGAC 93


RESULT    9
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
```

```
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match          16.1%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   1.0%;  Pred. No. 2.96e-04;
  Matches     1;  Conservative   31;  Mismatches  71;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnn 62
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|    :
Qy     22 GGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 81

Db     63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 105
          :  :  :  :  :  :     :  :  :  :  :  :  :  :
Qy     82 GCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTG 124


RESULT   10
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match          15.6%;  Score 28;  DB 12;  Length 114;
  Best Local Similarity   4.3%;  Pred. No. 1.00e-03;
  Matches     4;  Conservative   27;  Mismatches  63;  Indels   0;  Gaps   0;

Db      9 bnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnnbnnbtg 68
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      3 GGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTG 62

Db     69 cnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 102
          :  :     :  :  :  :  :  :  :  :  :
Qy     63 GCAGCTCACATGGAACAGGGCCGGGTATGACTTT 96


RESULT   11
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match          15.6%;  Score 28;  DB 12;  Length 114;
  Best Local Similarity   2.1%;  Pred. No. 1.00e-03;
  Matches     2;  Conservative   28;  Mismatches  64;  Indels   0;  Gaps   0;

Db      9 bnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnnbnnbnn 68
          :  :  :  :|| :  :  :  :.:  :  :  :  :  :  :  :  :  :  :
Qy      3 GGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTG 62

Db     69 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 102
          :  :  :     :  :  :  :  :  :  :
Qy     63 GCAGCTCACATGGAACAGGGCCGGGTATGACTTT 96
```

```
RESULT   12
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           15.6%;  Score 28;  DB 12;  Length 114;
  Best Local Similarity   2.8%;  Pred. No. 1.00e-03;
  Matches     3;  Conservative    30;  Mismatches  73;  Indels    0;  Gaps    0;

Db      3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :     :  :  :  :  :  :| |  :     :  :  :  :
Cp    107 GCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGA 48

Db     63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnb 108
          :  :  :  :  :  :  :  :|   :  :  :  :  :  :  :  :
Cp     47 GGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCT 2


RESULT   13
ID   T76405 standard; DNA; 178 BP.
AC   T76405;
DT   15-SEP-1997  (first entry)
DE   Human endothelin-1 antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Claim 5; Page 38; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human endothelin-1, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   178 BP;    0 A;    52 C;    46 G;    32 T;

  Query Match           15.6%;  Score 28;  DB 32;  Length 178;
  Best Local Similarity  25.0%;  Pred. No. 1.00e-03;
  Matches    18;  Conservative    32;  Mismatches  22;  Indels    0;  Gaps    0;

Db    100 cttggcbbgccbcbbbcbgcbgbgbgbbbbtcbtgbgcbbbtbbtccbttctgbbbbbbb 159
          || |||:: ||: :::|: |: : :| :::|||:| :|| ::|:    :     :::::::
Qy     27 CTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGG 86

Db    160 gggbtcbbbbbc 171
          |  :  :::::|
Qy     87 GTATGACTTTGC 98


RESULT   14
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
```

```
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           15.0%;  Score 27;  DB 12;  Length 114;
  Best Local Similarity   4.4%;  Pred. No. 3.36e-03;
  Matches     4;  Conservative   26;  Mismatches  61;  Indels   0;  Gaps   0;

Db     24 bnnbnnbnnbtgcnnbnnbnnbnnbnnbnnnnnnnnbnnbnnbnnbnnbnnbnnbtg 83
          :  :  :  : ||  :  :  :  :  :  :        :  :  :  :  :  :  :||
Qy      3 GGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTG 62

Db     84 cnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
          :  :     :  :  :  :  :  :  :
Qy     63 GCAGCTCACATGGAACAGGGCCGGGTATGAC 93


RESULT   15
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           15.0%;  Score 27;  DB 12;  Length 114;
  Best Local Similarity   3.7%;  Pred. No. 3.36e-03;
  Matches     4;  Conservative   29;  Mismatches  76;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnn 62
          :  :  :  :  :     |   :  :  :  :  :  :  :  :  :  :  :
Cp    110 TCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCA 51

Db     63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnb 111
             :  :  :  :  :  :  :  :  :  :  :  :|||  :  :
Cp     50 AGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCT 2


Search completed: Sun Oct 24 17:33:41 1999
Job time : 64 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:33:58 1999; MasPar time 16.98 Seconds
916.867 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-3
Description: (1-180) from US09092296.seq
Perfect Score: 180
N.A. Sequence: 1 CAGGAGCGCAGTGGCCACTA.....CTTGAAAAGCTCTGCCTCCT 180
Comp: GTCCTCGCGTCACCGGTGAT.....GAACTTTTCGAGACGGAGGA

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-issued
1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 7.135; Variance 3.967; scale 1.799

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description	Pred. No.
1		29	16.1	965	3	US-08-388-	Sequence 22, Applicati	6.29e-06
2		28	15.6	7218	2	US-08-232-	Sequence 14, Applicati	2.53e-05
3		27	15.0	215	1	US-08-238-	Sequence 5, Applicatio	1.01e-04
c	4	27	15.0	215	1	US-08-238-	Sequence 5, Applicatio	1.01e-04
c	5	26	14.4	965	3	US-08-388-	Sequence 22, Applicati	3.93e-04
6		22	12.2	75	4	PCT-US95-1	Sequence 99, Applicati	7.70e-02
7		22	12.2	82	4	PCT-US95-1	Sequence 97, Applicati	7.70e-02
c	8	22	12.2	3088	3	US-08-418-	Sequence 1, Applicatio	7.70e-02
c	9	21	11.7	65	1	US-08-471-	Sequence 145, Applicat	2.73e-01
10		21	11.7	66	1	US-08-471-	Sequence 144, Applicat	2.73e-01
c	11	21	11.7	68	1	US-08-471-	Sequence 143, Applicat	2.73e-01
12		21	11.7	69	1	US-08-471-	Sequence 142, Applicat	2.73e-01
13		21	11.7	74	4	PCT-US95-1	Sequence 94, Applicati	2.73e-01
c	14	21	11.7	74	4	PCT-US95-1	Sequence 100, Applicat	2.73e-01
c	15	21	11.7	75	4	PCT-US95-1	Sequence 99, Applicati	2.73e-01
16		21	11.7	81	4	PCT-US95-1	Sequence 98, Applicati	2.73e-01
17		21	11.7	81	4	PCT-US95-1	Sequence 92, Applicati	2.73e-01
c	18	21	11.7	81	4	PCT-US95-1	Sequence 98, Applicati	2.73e-01
c	19	21	11.7	82	4	PCT-US95-1	Sequence 97, Applicati	2.73e-01
20		21	11.7	906	3	US-09-031-	Sequence 41, Applicati	2.73e-01
21		21	11.7	906	3	US-08-847-	Sequence 41, Applicati	2.73e-01
c	22	21	11.7	906	3	US-08-847-	Sequence 40, Applicati	2.73e-01
c	23	21	11.7	906	3	US-09-031-	Sequence 40, Applicati	2.73e-01
24		21	11.7	908	3	US-09-031-	Sequence 39, Applicati	2.73e-01
25		21	11.7	908	3	US-08-847-	Sequence 39, Applicati	2.73e-01
c	26	21	11.7	908	3	US-08-847-	Sequence 37, Applicati	2.73e-01
c	27	21	11.7	908	3	US-09-031-	Sequence 37, Applicati	2.73e-01
28		21	11.7	909	3	US-08-847-	Sequence 26, Applicati	2.73e-01
29		21	11.7	909	3	US-09-031-	Sequence 26, Applicati	2.73e-01
c	30	21	11.7	909	3	US-09-031-	Sequence 25, Applicati	2.73e-01
c	31	21	11.7	909	3	US-08-847-	Sequence 25, Applicati	2.73e-01
32		21	11.7	911	3	US-09-031-	Sequence 24, Applicati	2.73e-01
33		21	11.7	911	3	US-08-847-	Sequence 24, Applicati	2.73e-01
c	34	21	11.7	911	3	US-08-847-	Sequence 22, Applicati	2.73e-01
c	35	21	11.7	911	3	US-09-031-	Sequence 22, Applicati	2.73e-01
36		21	11.7	5235	3	US-09-031-	Sequence 36, Applicati	2.73e-01
37		21	11.7	5235	3	US-08-847-	Sequence 36, Applicati	2.73e-01
c	38	21	11.7	5235	3	US-09-031-	Sequence 35, Applicati	2.73e-01
c	39	21	11.7	5235	3	US-08-847-	Sequence 35, Applicati	2.73e-01
40		21	11.7	5503	3	US-09-031-	Sequence 34, Applicati	2.73e-01
41		21	11.7	5503	3	US-08-847-	Sequence 34, Applicati	2.73e-01
c	42	21	11.7	5503	3	US-08-847-	Sequence 32, Applicati	2.73e-01
c	43	21	11.7	5503	3	US-09-031-	Sequence 32, Applicati	2.73e-01
c	44	20	11.1	66	4	PCT-US95-1	Sequence 93, Applicati	9.46e-01
c	45	20	11.1	2334	1	US-08-406-	Sequence 1, Applicatio	9.46e-01

ALIGNMENTS

RESULT 1
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC xxxxxx
DT
DE Sequence 22, Application US/08388672A
CC Sequence 22, Application US/08388672A
CC Patent No. 5795961
CC GENERAL INFORMATION:
CC APPLICANT: Wallace, T. Paul
CC APPLICANT: Harris, William J.
CC APPLICANT: Carr, Frank J.
CC APPLICANT: Old, Lloyd J.
CC APPLICANT: Welt, Sydney
CC APPLICANT: Kitamura, Kunio
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B
CC TITLE OF INVENTION: Antibodies
CC NUMBER OF SEQUENCES: 25
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Felfe and Lynch
CC STREET: 805 Third Avenue
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10022
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/388,672A
CC FILING DATE: 14-FEB-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hanson, No. 5795961man D.
CC REGISTRATION NUMBER: 30,946
CC REFERENCE/DOCKET NUMBER: LUD 5409
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-688-9200
CC TELEFAX: 212-838-3884
CC INFORMATION FOR SEQ ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 965 base pairs
CC TYPE: nucleic acid

```
CC       STRANDEDNESS:  unknown
CC       TOPOLOGY:  unknown
CC     MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

  Query Match           16.1%;  Score 29;  DB 3;  Length 965;
  Best Local Similarity  19.6%;  Pred. No. 6.29e-06;
  Matches    21;  Conservative   48;  Mismatches  37;  Indels   1;  Gaps   1;

Db    857 VRGRSYDSDGGDYWGGTTVTVSSHUVKDMTSSSSASVGDRVTTCRSSTTHGNGNTYYWYK 916
          ::| ::::|| ::   |: :: ::::::|::::   |: :| | : | :    |:::::
Qy      4 GAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACC-CTCCTTG 62

Db    917 GKAKYRVSNRSGVSRSGSGSGTDYTTSSDATYYCGTHARTGGTKVKG 963
          | |:: :  : |: :: :|:  :  |  :| :: | :| ||  :::|
Qy     63 GCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG 109


RESULT    2
ID   US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC   xxxxxx
DT
DE   Sequence 14, Application US/08232463
CC   Sequence 14, Application US/08232463
CC   Patent No. 5670367
CC    GENERAL INFORMATION:
CC      APPLICANT:  DORNER, F.
CC      APPLICANT:  SCHEIFLINGER, F.
CC      APPLICANT:  FALKNER, F. G.
CC      TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
CC      NUMBER OF SEQUENCES:  52
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Foley & Lardner
CC        STREET:  1800 Diagonal Road, Suite 500
CC        CITY:  Alexandria
CC        STATE:  VA
CC        COUNTRY:  USA
CC        ZIP:  22313-0299
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/232,463
CC        FILING DATE:
CC        CLASSIFICATION:  435
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  US/07/935,313
CC        FILING DATE:
CC        APPLICATION NUMBER:  EP 91 114 300.6
CC        FILING DATE:  26-AUG-1991
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  BENT, Stephen A.
CC        REGISTRATION NUMBER:  29,768
CC        REFERENCE/DOCKET NUMBER:  30472/114 IMMU
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703)836-9300
CC        TELEFAX:  (703)683-4109
CC        TELEX:  899149
CC     INFORMATION FOR SEQ ID NO:  14:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  7218 base pairs
CC         TYPE:  nucleic acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         CLONE:  pTZgpt-Fls
SQ   SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

  Query Match           15.6%;  Score 28;  DB 2;  Length 7218;
  Best Local Similarity   2.5%;  Pred. No. 2.53e-05;
  Matches     4;  Conservative   90;  Mismatches  66;  Indels   0;  Gaps   0;

Db   1061 TGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1120
          || | |::   :: :::::: ::::::::::  :::::::  :  ::: : :    :
Qy     21 TGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAG 80

Db   1121 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1180
             ::   : :  :::: :  ::    ::       ::::::::  :   :::::::: :
Qy     81 GGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAGTTCCTCCTATG 140

Db   1181 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1220
             :::  ::::::    :: :::      :::: ::::::
Qy    141 AGTCCAGCTTCCTGGAATTGCTTGAAAAGCTCTGCCTCCT 180


RESULT    3
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEA
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  215 base pairs
CC         TYPE:  nucleic acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       FEATURE:
CC         NAME/KEY:  misc_feature
CC         LOCATION:  1..215
CC         OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC         OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           15.0%;  Score 27;  DB 1;  Length 215;
  Best Local Similarity  20.0%;  Pred. No. 1.01e-04;
  Matches    22;  Conservative   42;  Mismatches  45;  Indels   1;  Gaps   1;

Db     18 CNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAK 77
          | :: : :| |  ::| :   :  |:|: :| :::  : |:  : :|   :|   :|| :
Qy     58 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG-AAGGAGT 116
```

```
Db      78 THYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
           :::|:| :: |: : :| ::   :    :: | |:   :||: :|: : |:
Qy     117 CTTTTCTGACAAGTTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAA 166


RESULT    4
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC     INFORMATION FOR SEQ ID NO:  5:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  215 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  protein
CC      FEATURE:
CC        NAME/KEY:  misc_feature
CC        LOCATION:  1..215
CC        OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC        OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           15.0%;  Score 27;  DB 1;  Length 215;
  Best Local Similarity  13.8%;  Pred. No. 1.01e-04;
  Matches    22;  Conservative   66;  Mismatches  69;  Indels   2;  Gaps   2;

Db       1 MTNVTMSSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNK 60
           :| : :::  :: : |:  ::|:::|   : : |    : :|:|:  ||:::  :: : :
Cp     158 ATTCCAGGAAGCTGGACTCATAGGAGGAACTTGTCAGAAAAGACTCCTTCAGCTTC-AGT 100

Db      61 YSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTD-GNRSGADS 119
           :::|   :    :   : ::: :| :   ::  :|::: |:| :::    : | ::||
Cp      99 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 40

Db     120 YGSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTK 158
           |::: |  :  :       |:::: :   ::|::  |:
Cp      39 AGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTG 1


RESULT    5
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
CC        TELEFAX:  212-838-3884
CC     INFORMATION FOR SEQ ID NO:  22:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  965 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  unknown
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

  Query Match           14.4%;  Score 26;  DB 3;  Length 965;
  Best Local Similarity  17.8%;  Pred. No. 3.93e-04;
  Matches    16;  Conservative   43;  Mismatches  30;  Indels   1;  Gaps   1;

Db     781 GURHUVHVSGGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKRGRVTMADTSS 840
           |: ::::::||: :|:| | ::: |:   ::  :  |: || :    |::|:: :|: ::
Cp      93 GTCATACCCGGCCCTGT-TCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 35

Db     841 NSRSSVTAADTAVYYCVRGRSYDSDGGDYW 870
           ::::  |   : : : |:::::| :::
Cp      34 CAGCCCAGACCCCATAGTGGCCACTGCGCT 5


RESULT    6
ID   PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.
AC   xxxxxx
DT
DE   Sequence 99, Application PC/TUS9511934
CC   Sequence 99, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
```

```
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  99:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  75 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 75 BP; 1 A; 1 C; 7 G; 5 T; 61 OTHER.

  Query Match           12.2%;  Score 22;  DB 4;  Length 75;
  Best Local Similarity   8.7%;  Pred. No. 7.70e-02;
  Matches     6;  Conservative   19;  Mismatches  44;  Indels   0;  Gaps   0;

Db       5 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 64
           |   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      23 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 82

Db      65 NBGGTTGTG 73
             :|| | ||
Qy      83 CCGGGTATG 91


RESULT    7
ID   PCT-US95-11934-97 STANDARD; DNA; UNC; 82 BP.
AC   xxxxxx
DT
DE   Sequence 97, Application PC/TUS9511934
CC   Sequence 97, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  97:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  82 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 82 BP; 1 A; 2 C; 10 G; 8 T; 61 OTHER.

  Query Match           12.2%;  Score 22;  DB 4;  Length 82;
  Best Local Similarity   8.7%;  Pred. No. 7.70e-02;
  Matches     6;  Conservative   19;  Mismatches  44;  Indels   0;  Gaps   0;

Db      12 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 71
           |   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      23 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 82

Db      72 NBGGTTGTG 80
             :|| | ||
Qy      83 CCGGGTATG 91


RESULT    8
ID   US-08-418-444A-1 STANDARD; DNA; UNC; 3088 BP.
AC   xxxxxx
DT
DE   Sequence 1, Application US/08418444A
CC   Sequence 1, Application US/08418444A
CC   Patent No. 5773688
CC    GENERAL INFORMATION:
CC      APPLICANT:  KURODA, HISAO
CC      APPLICANT:  HIROTA, NAOHIKO
CC      APPLICANT:  ITO, KAZUTOSHI
CC      TITLE OF INVENTION:  GENE EXPRESSION VECTOR USING THE GENE
CC      TITLE OF INVENTION:  EXPRESSION REGULATING REGION OF THE ADP RIBOSYLATION
CC      TITLE OF INVENTION:  FACTOR
CC      NUMBER OF SEQUENCES:  9
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
CC        STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC        CITY:  ARLINGTON
CC        STATE:  VIRGINIA
CC        COUNTRY:  USA
CC        ZIP:  22202
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/418,444A
CC        FILING DATE:  07-APR-1995
CC        CLASSIFICATION:  800
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  JP HEI 6-71048
CC        FILING DATE:  08-APR-1994
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Oblon, No. 5773688man F.
CC        REGISTRATION NUMBER:  24,618
CC        REFERENCE/DOCKET NUMBER:  2589-024-0
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703) 413-3000
CC        TELEFAX:  (703) 413-2220
```

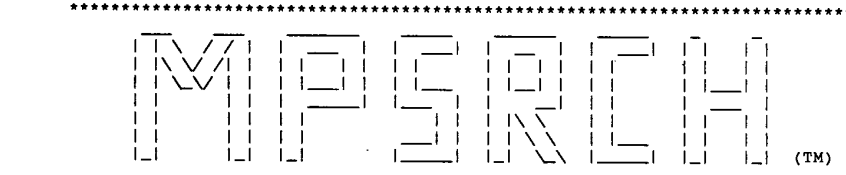
```
CC        TELEX:  248855 OPAT UR
CC    INFORMATION FOR SEQ ID NO:  1:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  3088 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  double
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 3088 BP; 716 A; 761 C; 672 G; 939 T; 0 OTHER.

  Query Match           12.2%;  Score 22;  DB 3;  Length 3088;
  Best Local Similarity  72.0%;  Pred. No. 7.70e-02;
  Matches    36;  Conservative    0;  Mismatches  14;  Indels   0;  Gaps   0;

Db     15 CCGAAGATGGGCCAGAGGGGGGCCAGGGGCTGCCCAGGCGCCCTGCTGGC 64
           || | || || | ||||| || | |||||| |||||| | || |  ||||
Cp     63 CCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGC 14


RESULT    9
ID   US-08-471-052A-145 STANDARD; DNA; UNC; 65 BP.
AC   xxxxxx
DT
DE   Sequence 145, Application US/08471052A
CC   Sequence 145, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  145:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  65 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 65 BP; 3 A; 3 C; 3 G; 2 T; 54 OTHER.

  Query Match           11.7%;  Score 21;  DB 1;  Length 65;
  Best Local Similarity  13.8%;  Pred. No. 2.73e-01;
  Matches     9;  Conservative   16;  Mismatches  40;  Indels   0;  Gaps   0;

Db      1 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNA 60
          || ||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  |
Cp    147 CTGGACTCATAGGAGGAACTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATA 88

Db     61 CCTGG 65
```

```
          || ||
Cp     87 CCCGG 83


RESULT   10
ID   US-08-471-052A-144 STANDARD; DNA; UNC; 66 BP.
AC   xxxxxx
DT
DE   Sequence 144, Application US/08471052A
CC   Sequence 144, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  144:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  66 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 66 BP; 2 A; 3 C; 4 G; 2 T; 55 OTHER.

  Query Match           11.7%;  Score 21;  DB 1;  Length 66;
  Best Local Similarity   5.2%;  Pred. No. 2.73e-01;
  Matches     3;  Conservative   18;  Mismatches  37;  Indels   0;  Gaps   0;

Db      4 AGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBC 61
          ||   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy     10 AGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGC 67


RESULT   11
ID   US-08-471-052A-143 STANDARD; DNA; UNC; 68 BP.
AC   xxxxxx
DT
DE   Sequence 143, Application US/08471052A
CC   Sequence 143, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
```

```
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036-2711
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/471,052A
CC         FILING DATE:  06-JUNE-1995
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Misrock, S. Leslie
CC         REGISTRATION NUMBER:  18,872
CC         REFERENCE/DOCKET NUMBER:  1101-179
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212 790-9090
CC         TELEFAX:  212 869-8864/9741
CC         TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  143:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  68 bases
CC         TYPE:  nucleic acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  DNA
SQ   SEQUENCE 68 BP; 3 A; 3 C; 5 G; 3 T; 54 OTHER.

  Query Match           11.7%;  Score 21;  DB 1;  Length 68;
  Best Local Similarity  13.8%;  Pred. No. 2.73e-01;
  Matches     9;  Conservative   16;  Mismatches  40;  Indels   0;  Gaps   0;

Db       4 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNA 63
           || ||:  :  :  :  :     :  :  :  :  :  :  :  :  :     :  :  :  |
Cp     147 CTGGACTCATAGGAGGAACTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATA 88

Db      64 CCTGG 68
           || ||
Cp      87 CCCGG 83


RESULT   12
ID   US-08-471-052A-142 STANDARD; DNA; UNC; 69 BP.
AC   xxxxxx
DT
DE   Sequence 142, Application US/08471052A
CC   Sequence 142, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  142:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  69 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 69 BP; 2 A; 4 C; 6 G; 2 T; 55 OTHER.

  Query Match           11.7%;  Score 21;  DB 1;  Length 69;
  Best Local Similarity  5.2%;  Pred. No. 2.73e-01;
  Matches     3;  Conservative   18;  Mismatches  37;  Indels   0;  Gaps   0;

Db       7 AGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBC 64
           ||   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy      10 AGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGC 67


RESULT   13
ID   PCT-US95-11934-94 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 94, Application PC/TUS9511934
CC   Sequence 94, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  94:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 3 A; 4 C; 3 G; 1 T; 63 OTHER.

  Query Match           11.7%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity  1.7%;  Pred. No. 2.73e-01;
  Matches     1;  Conservative   20;  Mismatches  39;  Indels   0;  Gaps   0;

Db       9 BNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNA 68
```

```
                  : : : : : : : : : : : : : : : : : : : : : : : : : : : |
Qy      6 GCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCA 65


RESULT   14
ID   PCT-US95-11934-100 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 100, Application PC/TUS9511934
CC   Sequence 100, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  100:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 6 A; 6 C; 1 G; 1 T; 60 OTHER.

  Query Match           11.7%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity   7.5%;  Pred. No. 2.73e-01;
  Matches     5;  Conservative   19;  Mismatches  43;  Indels   0;  Gaps   0;

Db       5 AVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVN 64
           |: : : : : : : :     : : : : : : : : : : : : : : : : : : :
Cp      94 AGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 35

Db      65 NACCACA 71
            | | ||
Cp      34 CAGCCCA 28


RESULT   15
ID   PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.
AC   xxxxxx
DT
DE   Sequence 99, Application PC/TUS9511934
CC   Sequence 99, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  99:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  75 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 75 BP; 1 A; 1 C; 7 G; 5 T; 61 OTHER.

  Query Match           11.7%;  Score 21;  DB 4;  Length 75;
  Best Local Similarity   7.6%;  Pred. No. 2.73e-01;
  Matches     5;  Conservative   18;  Mismatches  43;  Indels   0;  Gaps   0;

Db       4 AGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNB 63
           ||   : : : : : : :     : : : : : : : : : : :     : : :
Cp     118 AGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAG 59

Db      64 NNBGGT 69
              :|||
Cp      58 GAGGGT 53


Search completed: Sun Oct 24 17:34:17 1999
Job time : 19 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:26:03 1999; MasPar time 365.62 Seconds
1153.579 Million cell updates/sec
Tabular output not generated.

Title: >US-09-092-296-3
Description: (1-180) from US09092296.seq
Perfect Score: 180
N.A. Sequence: 1 CAGGAGCGCAGTGGCCACTA.....CTTGAAAAGCTCTGCCTCCT 180
Comp: GTCCTCGCGTCACCGGTGAT.....GAACTTTTCGAGACGGAGGA

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est2 20:gb_est20
21:gb_est21 22:gb_est22 23:gb_est23 24:gb_est24
25:gb_est25 26:gb_est26 27:gb_est27 28:gb_est28
29:gb_est29 30:gb_est3 31:gb_est4 32:gb_est5 33:gb_est6
34:gb_est7 35:gb_est8 36:gb_est9 37:gb_gss1 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 9.460; Variance 1.952; scale 4.848

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	51	28.3	252	17	AA754459	97SN1787 Rice Immature	9.68e-50
c	2	43	23.9	328	23	AI136523	UI-R-C2p-nq-e-02-0-UI.	3.33e-36
c	3	39	21.7	252	17	AA754459	97SN1787 Rice Immature	1.13e-29
c	4	31	17.2	247	17	AA754458	97SN1784 Rice Immature	2.88e-17
	5	30	16.7	247	17	AA754458	97SN1784 Rice Immature	8.54e-16
c	6	28	15.6	2275	20	AF034173	AF034173 Human mRNA (T	6.48e-13
	7	27	15.0	2275	20	AF034173	AF034173 Human mRNA (T	1.64e-11
	8	23	12.8	311	11	AA323964	EST26816 Cerebellum II	3.57e-06
c	9	23	12.8	459	13	AA475002	vh08h04.r1 Soares mous	3.57e-06
	10	23	12.8	1287	20	AF038250	AF038250 Human mRNA (T	3.57e-06
	11	22	12.2	259	38	B81136	CIT-HSP-2015F16.TFC CI	6.38e-05
	12	22	12.2	308	8	D40392	RICS2342A Rice shoot O	6.38e-05
	13	22	12.2	339	19	F08745	HSC1DB011 normalized i	6.38e-05
c	14	22	12.2	427	33	N80550	za10h06.r1 Soares feta	6.38e-05
	15	22	12.2	433	34	W79098	zd75h10.r1 Soares_feta	6.38e-05
	16	22	12.2	441	14	C28493	C28493 Rice callus cDN	6.38e-05
	17	22	12.2	513	40	AQ235722	HS_2015_B2_C08_T7 CIT	6.38e-05
c	18	22	12.2	634	22	AI063013	GH02423.5prime GH Dros	6.38e-05
	19	22	12.2	1025	37	B12587	F22O11-Sp6.1 IGF Arabi	6.38e-05
	20	21	11.7	243	8	D22101	RICC10342A Rice callus	1.04e-03
	21	21	11.7	288	11	AA335414	EST39832 Epididymus Ho	1.04e-03
c	22	21	11.7	299	26	AI382671	qz05f06.x1 NCI_CGAP_CL	1.04e-03
c	23	21	11.7	299	8	M79528	wEST00065 Mixed stage,	1.04e-03
	24	21	11.7	317	19	F10052	HSC39H122 normalized i	1.04e-03
	25	21	11.7	343	8	T47417	yb13f12.r1 Stratagene	1.04e-03
	26	21	11.7	364	34	W33870	mc56b03.r1 Soares mous	1.04e-03
c	27	21	11.7	412	38	AQ007423	CIT-HSP-2292F15.TR CIT	1.04e-03
	28	21	11.7	416	8	T08769	EST06661 Infant Brain,	1.04e-03
	29	21	11.7	417	30	R36249	yh91f09.r1 Soares plac	1.04e-03
c	30	21	11.7	422	10	AA275959	vc27e03.r1 Barstead MP	1.04e-03
	31	21	11.7	426	8	T15255	crs855 lambdaZAPST Ric	1.04e-03
c	32	21	11.7	437	37	FR0032574	Fugu rubripes GSS sequ	1.04e-03
	33	21	11.7	451	38	AQ024376	HS_2182_A1_D09_MF CIT	1.04e-03
	34	21	11.7	468	35	AA046996	zf50a10.r1 Soares reti	1.04e-03
c	35	21	11.7	475	28	AI522104	ti83g06.x1 NCI_CGAP_Ki	1.04e-03
c	36	21	11.7	502	15	AA586074	28723 Lambda-PRL2 Arab	1.04e-03
c	37	21	11.7	516	31	H41536	yp71c01.s1 Soares adul	1.04e-03
c	38	21	11.7	524	10	AA239967	mw24g07.r1 Soares mous	1.04e-03
c	39	21	11.7	533	15	AA593745	nl83g11.s1 NCI_CGAP_Br	1.04e-03
	40	21	11.7	557	23	AI102979	EST212268 Normalized r	1.04e-03
c	41	21	11.7	567	30	R61539	yh16f01.s1 Soares infa	1.04e-03
	42	21	11.7	630	39	AQ201160	RPCI11-46K18.TJ RPCI11	1.04e-03
c	43	21	11.7	694	36	AA140679	CK00462.5prime CK Dros	1.04e-03
	44	21	11.7	762	37	B19344	T27I1-Sp6.1 TAMU Arabi	1.04e-03
c	45	20	11.1	492	34	W96005	ze09b02.r1 Soares_feta	1.55e-02

ALIGNMENTS

RESULT 1
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID g2801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
FEATURES Location/Qualifiers
source 1. .252
/organism="Oryza sativa"

/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5'end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 28.3%; Score 51; DB 17; Length 252;
Best Local Similarity 12.3%; Pred. No. 9.68e-50;
Matches 20; Conservative 84; Mismatches 56; Indels 2; Gaps 2;

```
Db      29 BVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWBVBNTKVD 88
           :: : :|::| :: ::   ::||: : :    : :|:||: ::  : : :: : : |:::
Qy       4 GAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGG 63

Db      89 VGNHTRCSRWRBVTRMAH-YHDYTNCBBYNNNDYHMWHBBMYBBTGCMTCTMWCWBHYNT 147
           :  :|  :::::: :   :::    ::    : :::::::: ::|| :    :|:::: |
Qy      64 CAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCT 123

Db     148 KCTASGWHTSTNYDVKS-STNTWGVTBSYDKSMHGYWCSBBV 188
           :  | |:: :| ::::: :|  :|:|:::::::: : | :::
Qy     124 GACAAGTTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGA 165
```

RESULT 2
LOCUS AI136523 328 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C2p-nq-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone UI-R-C2p-nq-e-02-0-UI 3', mRNA sequence.
ACCESSION AI136523
NID g3637300
VERSION AI136523.1 GI:3637300
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 328)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 14, 1998 this sequence version replaced gi:1877567.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.
FEATURES Location/Qualifiers
source 1. .328
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
/db_xref="taxon:10116"
/clone="UI-R-C2p-nq-e-02-0-UI"
/clone_lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
BASE COUNT 62 a 77 c 98 g 91 t
ORIGIN

Query Match 23.9%; Score 43; DB 23; Length 328;
Best Local Similarity 73.6%; Pred. No. 3.33e-36;
Matches 67; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Db     210 AGGAGGCAGATCTTTTGGAGCATGTCCAGAAAGCCGGAGTCTTGGGAGGTCTTGGCCTGA 269
           |||||||||| |||||  ||||  ||||| |||| ||| || | |||||   | | | ||
Cp     180 AGGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGGAGGAACTTGTCAGA 121

Db     270 GATGCCTCGGTCAGTTTTACTTGCAAAGTCA 300
           | | |||  |||| || | ||||||||||
Cp     120 AAAGACTCCTTCAGCTTCAGTTGCAAAGTCA 90
```

RESULT 3
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID g2801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
FEATURES Location/Qualifiers
source 1. .252
/organism="Oryza sativa"

```
                /cultivar="Milyang23"
                /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                XhoI; Directional cDNA library inserted into lambda ZAPII
                vector  at 5'end with EcoRI and 3' end with Xho I site."
                /db_xref="taxon:4530"
                /map="6"
                /clone="97SN1787"
                /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                /tissue_type="Immature Seed"
                /dev_stage="5 days after pollination"
                /lab_host="E. coli SOLR"
BASE COUNT        5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match           21.7%;  Score 39;  DB 17;  Length 252;
  Best Local Similarity   8.8%;  Pred. No. 1.13e-29;
  Matches    10;  Conservative   61;  Mismatches  43;  Indels   0;  Gaps   0;

Db      68 VNWBNVSGDWHYWBVBNTKVDVGNHTRCSRWRBVTRMAHYHDYTNCBBYNNNDYHMWHBB 127
           : ::  :|: ::::::  ::::  :  | ::: : :: :::::  |:::   : ::::::
Cp     151 GAAGCTGGACTCATAGGAGGAACTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGT 92

Db     128 MYBBTGCMTCTMWCWBHYNTKCTASGWHTSTNYDVKSSTNTWGVTBSYDKSMHG 181
           : :   |  | :: :::: |:  |: : :   :::::|  :|: :  : :::|
Cp      91 CATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAG 38


RESULT    4
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
VERSION     AA754458.1  GI:2801164
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797455.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .247
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /map="6"
                      /clone="97SN1784"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                      /tissue_type="Immature Seed"
                      /dev_stage="5 days after pollination"
                      /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match           17.2%;  Score 31;  DB 17;  Length 247;
  Best Local Similarity   8.3%;  Pred. No. 2.88e-17;
  Matches    11;  Conservative   71;  Mismatches  49;  Indels   2;  Gaps   2;

Db     100 VNVHSGWNNRCSNSVVYVWBTAYCDYBHYBDRANHVDDTRCTNDRGYCNYTASDNGTSAT 159
           : ::: :   |  :::  :: |  : ::::::  :::: :   :: :  : |::   :|
Cp     174 CAGAGCTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGGAGGAACTTGTCAGAAAAGAC 115

Db     160 KRVTGYDKTDSDCGGGCWRKVTYGSSBYBRCGVNVMVRTTSMWTDKSTKMBSM-DMSRRS 218
           : :| ::: :   |    : : : : : : ||: :: :||:::|::: ::::: ::::::
Cp     114 TCCTTCAGCTTCAGTTGCAAAGT-CATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAG 56

Db     219 RVHYGRWMBNKKR 231
           :::: : ::  ::
Cp      55 GGTCAAGAGGAGG 43


RESULT    5
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
VERSION     AA754458.1  GI:2801164
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797455.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES              Location/Qualifiers
     source           1. .247
                      /organism="Oryza sativa"
                      /cultivar="Milyang23"
                      /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                      XhoI; Directional cDNA library inserted into lambda ZAPII
                      vector  at 5'end with EcoRI and 3' end with Xho I site."
                      /db_xref="taxon:4530"
                      /map="6"
                      /clone="97SN1784"
                      /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                      /tissue_type="Immature Seed"
                      /dev_stage="5 days after pollination"
                      /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match           16.7%;  Score 30;  DB 17;  Length 247;
  Best Local Similarity   9.8%;  Pred. No. 8.54e-16;
  Matches    13;  Conservative   64;  Mismatches  55;  Indels   0;  Gaps   0;

Db     111 SNSVVYVWBTAYCDYBHYBDRANHVDDTRCTNDRGYCNYTASDNGTSATKRVTGYDKTDS 170
           : :::::::  :|:::::::::| :::       :|:    | :   :| ::: | :: ::
```

```
Qy      32 CTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATG 91

Db     171 DCGGGCWRKVTYGSSBYBRCGVNVMVRTTSMWTDKSTKMBSMDMSRRSRVHYGRWMBNKK 230
           :|     : :|   ::::: |: : : || :: :   : :::::: : :: :  : : :
Qy      92 ACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAGTTCCTCCTATGAGTCCAGCTTC 151

Db     231 RGMSRNWTDTKT 242
             :: :|: :|
Qy     152 CTGGAATTGCTT 163


RESULT    6
LOCUS       AF034173     2275 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
VERSION     AF034173.1  GI:2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045115.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match           15.6%;  Score 28;  DB 20;  Length 2275;
  Best Local Similarity  12.3%;  Pred. No. 6.48e-13;
  Matches     9;  Conservative   44;  Mismatches  19;  Indels   1;  Gaps   1;

Db   1523 RGMKKCWKYRKYKKYTSTYYKSWSRWYWYTTYTYWYCWCCT-SMKSASCAMMRWMGYMGS 1581
          :|: :|:: :: ::  :|: :::::: :   : :::|: || :::  :||::::: :: :
Cp    145 GGACTCATAGGAGGAACTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACC 86

Db   1582 RSSRSYWGYWGSM 1594
           :: :::|:: ::
Cp     85 CGGCCCTGTTCCA 73


RESULT    7
LOCUS       AF034173     2275 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
VERSION     AF034173.1  GI:2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045115.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match           15.0%;  Score 27;  DB 20;  Length 2275;
  Best Local Similarity  17.8%;  Pred. No. 1.64e-11;
  Matches    19;  Conservative   53;  Mismatches  33;  Indels   2;  Gaps   2;

Db   1534 YKKYTSTYYKSWSRWYWYTTYTYWYCWCCTSMKSASCAMMRWMGYMGSRSSRSYWGYWGS 1593
          :::: :|::::: ::: :| :|:  |: ||::   : |:::  |::|::   :   ::|:
Qy     39 TTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGC 98

Db   1594 MSGCYGMTKRYYRYSWTGWTKWTTWWYMWSMTRWTMTTYTTWWTWRT 1640
           :  |:|: : : : :  |: ::|| : : :: ::|: |: |:: ::|
Qy     99 AA-CTGAAGCTGAAGGAGTCTTTTCTGA-CAAGTTCCTCCTATGAGT 143


RESULT    8
LOCUS       AA323964     311 bp    mRNA            EST       20-APR-1997
DEFINITION  EST26816 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA323964
NID         g1976290
VERSION     AA323964.1  GI:1976290
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693635.

            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
```

```
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .311
                      /organism="Homo sapiens"
                      /note="Organ: brain; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):124516"
                      /db_xref="taxon:9606"
                      /clone_lib="Cerebellum II"
                      /tissue_type="cerebellum"
                      /dev_stage="adult"
BASE COUNT       59 a    110 c     74 g     64 t      4 others
ORIGIN

  Query Match           12.8%;  Score 23;  DB 11;  Length 311;
  Best Local Similarity  89.7%;  Pred. No. 3.57e-06;
  Matches    26;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db       8 TGGCCACTGGGGGGCCTGGGCTGCCCCTT 36
           ||||||||  |||| ||||||||||||||
Qy      12 TGGCCACTATGGGGTCTGGGCTGCCCCTT 40


RESULT    9
LOCUS       AA475002      459 bp    mRNA            EST       18-JUN-1997
DEFINITION  vh08h04.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:874903 5' similar to TR:G466574 G466574 THERMOSTABLE DNA
            POLYMERASE I ;, mRNA sequence.
ACCESSION   AA475002
NID         g2200761
VERSION     AA475002.1  GI:2200761
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 25, 1995 this sequence version replaced gi:637724.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:514383
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 430.
FEATURES              Location/Qualifiers
     source           1. .459
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Organ: mammary gland; Vector: pT7T3D-Pac
                      (Pharmacia) with a modified polylinker; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
                      /db_xref="taxon:10090"
                      /clone="IMAGE:874903"
                      /clone_lib="Soares mouse mammary gland NbMMG"
                      /sex="male"
                      /tissue_type="mammary gland"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
BASE COUNT      124 a    111 c    101 g    123 t
ORIGIN

  Query Match           12.8%;  Score 23;  DB 13;  Length 459;
  Best Local Similarity  82.9%;  Pred. No. 3.57e-06;
  Matches    29;  Conservative    0;  Mismatches   6;  Indels   0;  Gaps   0;

Db     306 GGATCCAGAGCTTTTGAAGTTATTCCAGGAATCTG 340
           |||  |||||||||| |||  |||||||||| |||
Cp     179 GGAGGCAGAGCTTTTCAAGCAATTCCAGGAAGCTG 145


RESULT   10
LOCUS       AF038250      1287 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF038250 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon9, mRNA sequence.
ACCESSION   AF038250
NID         g2815880
VERSION     AF038250.1  GI:2815880
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1287)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045085.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES              Location/Qualifiers
     source           1. .1287
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="6p21.3"
                      /clone="ntcon9"
                      /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      349 a    219 c    293 g    361 t     65 others
ORIGIN

  Query Match           12.8%;  Score 23;  DB 20;  Length 1287;
  Best Local Similarity  25.8%;  Pred. No. 3.57e-06;
  Matches    24;  Conservative   32;  Mismatches  37;  Indels   0;  Gaps   0;

Db     381 VVVATCNAGSCCVACVTCBCDCBTGGGDTBSHBVCMCBCBANADGATBATCGKVGGVBGA 440
           ::: ||  |:| : |:  : :|:|    |:::::| |: :   :| |:|   ::  : |
Qy      22 GGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 81

Db     441 STCMTHYNCCDCKTCGSAGVTVNHHDWSMAGGA 473
           : |   :   :|:| |:| :|: :  :::||||
Qy      82 GCCGGGTATGACTTTGCAACTGAAGCTGAAGGA 114
```

```
RESULT   11
LOCUS       B81136        259 bp    DNA             GSS       24-JUN-1998
DEFINITION  CIT-HSP-2015F16.TFC CIT-HSP Homo sapiens genomic clone 2015F16,
            genomic survey sequence.
ACCESSION   B81136
NID         g2868159
VERSION     B81136.1  GI:2868159
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 259)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .259
                     /organism="Homo sapiens"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
                     /db_xref="taxon:9606"
                     /clone="2015F16"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
BASE COUNT       68 a     43 c     56 g     92 t
ORIGIN

  Query Match           12.2%;  Score 22;  DB 38;  Length 259;
  Best Local Similarity  89.3%;  Pred. No. 6.38e-05;
  Matches    25;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db      59 GCTTCCTGGAATTTCTTGAAAATCGCTG 86
           ||||||||||||| |||||||| | |||
Qy     147 GCTTCCTGGAATTGCTTGAAAAGCTCTG 174


RESULT   12
LOCUS       D40392        308 bp    mRNA            EST       11-NOV-1994
DEFINITION  RICS2342A Rice shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D40392
NID         g569543
VERSION     D40392.1  GI:569543
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 308)
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .308
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /note="Etiolated shoot (8 days old)"
                     /db_xref="taxon:4530"
                     /clone_lib="Rice shoot"
BASE COUNT       57 a    126 c     57 g     68 t
ORIGIN

  Query Match           12.2%;  Score 22;  DB 8;  Length 308;
  Best Local Similarity  73.9%;  Pred. No. 6.38e-05;
  Matches    34;  Conservative    0;  Mismatches  12;  Indels   0;  Gaps   0;

Db     115 GCTGCACCTCCTCCTCCTCCTCCTCCTCCTCAGGAGCTCCAATGGA 160
           ||||| |||  |||||||| |   ||||||  | |||||  |||||
Qy      31 GCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGA 76


RESULT   13
LOCUS       F08745        339 bp    mRNA            EST       20-FEB-1995
DEFINITION  HSC1DB011 normalized infant brain cDNA Homo sapiens cDNA clone
            c-1db01, mRNA sequence.
ACCESSION   F08745
NID         g672165
VERSION     F08745.1  GI:672165
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     On Sep 21, 1992 this sequence version replaced gi:276248.

            Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y3c-1db01
            Seq primer: (-21)M13_universal
            High quality sequence stop: 277.
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total
                     brain;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
                     /db_xref="taxon:9606"
                     /clone="c-1db01"
```

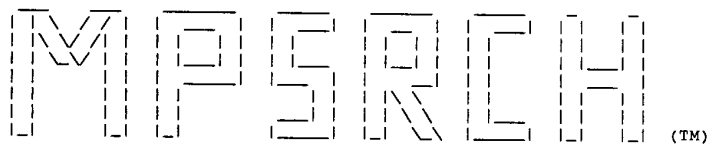
```
                 /clone_lib="normalized infant brain cDNA"
                 /sex="Female"
                 /tissue_type="total brain"
                 /dev_stage="3 months old"
BASE COUNT       57 a    116 c    100 g     64 t      2 others
ORIGIN

  Query Match           12.2%;  Score 22;  DB 19;  Length 339;
  Best Local Similarity  66.2%;  Pred. No. 6.38e-05;
  Matches    43;  Conservative    0;  Mismatches  22;  Indels   0;  Gaps   0;

Db      81 TGCCCCCTCCGGAAGCTTGGATGCCCCTNCACACCCTCTTGATCTTCCCTGTGATGTCAC 140
           || || ||  ||   || || |||||||   |  |||||||| | ||| ||  |  ||||
Qy      12 TGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCAC 71

Db     141 CTGGA 145
            ||||
Qy      72 ATGGA 76


RESULT   14
LOCUS       N80550        427 bp    mRNA            EST       02-APR-1996
DEFINITION  za10h06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:292187 5', mRNA sequence.
ACCESSION   N80550
NID         g1243251
VERSION     N80550.1  GI:1243251
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: reverse ET
            High quality sequence stop: 153.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Homo sapiens"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
                     /db_xref="GDB:3799961"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:292187"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      106 a    123 c    107 g     84 t      7 others
ORIGIN
```

```
  Query Match           12.2%;  Score 22;  DB 33;  Length 427;
  Best Local Similarity  70.8%;  Pred. No. 6.38e-05;
  Matches    34;  Conservative    0;  Mismatches  14;  Indels   0;  Gaps   0;

Db     251 AACTTGATGGAAAAGANTGCTTCAGCCTTNGTTGGACCGGCAAGCCCG 298
           ||||||   ||||||| | ||||||| |  |||| |  | ||  ||||
Cp     131 AACTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCG 84


RESULT   15
LOCUS       W79098        433 bp    mRNA            EST       17-OCT-1996
DEFINITION  zd75h10.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:346531 5', mRNA sequence.
ACCESSION   W79098
NID         g1390534
VERSION     W79098.1  GI:1390534
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 433)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 8, 1995 this sequence version replaced gi:801256.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 826    Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 379.
FEATURES             Location/Qualifiers
     source          1. .433
                     /organism="Homo sapiens"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
                     /db_xref="GDB:1271906"
                     /db_xref="taxon:9606"
                     /map="12"
                     /clone="IMAGE:346531"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       66 a    135 c    157 g     68 t      7 others
ORIGIN

  Query Match           12.2%;  Score 22;  DB 34;  Length 433;
  Best Local Similarity  75.0%;  Pred. No. 6.38e-05;
  Matches    30;  Conservative    0;  Mismatches  10;  Indels   0;  Gaps   0;

Db     317 CCCCTTCCCTCCTCCAGACCCTCCTTGGCCGGTGACNTNG 356
```

```
          ||| |  |||||||  |||||||||||||| | | || | |
Qy      36 CCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGG 75
```

Search completed: Sun Oct 24 17:32:17 1999
Job time : 374 secs.

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 18:41:27 1999; MasPar time 5.65 Seconds
293.555 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect Score: 558
Sequence: 1 MGSGLPLVLLLTLLGSSHGT.....SGTSVTLHHARSQHHVVCNT 78

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 27.329; Variance 130.705; scale 0.209

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	558	100.0	78	39	W88498	Human stomach carcino	2.84e-38
2	86	15.4	46	36	W70327	Secreted protein FB78	3.89e+01
3	86	15.4	47	38	W73408	Human secreted protei	3.89e+01
4	85	15.2	325	19	W02702	G-protein coupled bov	4.56e+01
5	85	15.2	325	16	R48730	G-protein coupled bov	4.56e+01
6	83	14.9	194	9	R47338	Peptide fragment of t	6.23e+01
7	83	14.9	411	31	W57046	Mouse apoptosis induc	6.23e+01
8	82	14.7	110	5	R26954	Human T lymphocyte re	7.28e+01
9	82	14.7	169	23	W21674	Human mitochondrial e	7.28e+01
10	81	14.5	401	35	W59924	Human 7-transmembrane	8.50e+01
11	81	14.5	1253	25	W10038	Mad binding protein,	8.50e+01
12	81	14.5	1261	25	W10040	Mad binding protein,	8.50e+01
13	80	14.3	398	3	R15138	Human serotonin 1D re	9.93e+01
14	80	14.3	713	2	R10052	Cyclomaltodextrin glu	9.93e+01
15	80	14.3	713	1	R06110	Sequence of cyclomalt	9.93e+01
16	80	14.3	945	12	R66060	Human NMDAR2 receptor	9.93e+01
17	80	14.3	1214	38	W87509	Human N-methyl-D-aspa	9.93e+01
18	80	14.3	1214	12	R66065	Human NMDAR2 receptor	9.93e+01
19	80	14.3	1219	12	R66063	Human NMDAR2 receptor	9.93e+01
20	80	14.3	1219	38	W87507	Human N-methyl-D-aspa	9.93e+01
21	80	14.3	1231	38	W87506	Human N-methyl-D-aspa	9.93e+01
22	80	14.3	1231	12	R66062	Human NMDAR2 receptor	9.93e+01
23	80	14.3	1236	38	W85574	Human N-methyl-D-aspa	9.93e+01
24	80	14.3	1236	12	R66037	Human N-methyl-D-aspa	9.93e+01
25	80	14.3	1239	38	W87508	Human N-methyl-D-aspa	9.93e+01
26	80	14.3	1239	12	R66064	Human NMDAR2 receptor	9.93e+01
27	80	14.3	1244	38	W87505	Human N-methyl-D-aspa	9.93e+01
28	80	14.3	1244	12	R66061	Human NMDAR2 receptor	9.93e+01
29	79	14.2	620	22	W14993	Human c-Fos induced g	1.16e+02
30	79	14.2	652	34	W49879	Amino acid sequence o	1.16e+02
31	79	14.2	969	24	W25170	Human insulinoma-asso	1.16e+02
32	79	14.2	986	24	W25171	Human insulinoma-asso	1.16e+02
33	79	14.2	1015	23	W18092	Type I diabetes-assoc	1.16e+02
34	79	14.2	1015	27	W35345	Human protein tyrosin	1.16e+02
35	78	14.0	99	39	W67828	Human secreted protei	1.35e+02
36	78	14.0	120	17	R93614	Kaposi's sarcoma asso	1.35e+02
37	78	14.0	120	18	R97838	Kaposi's sarcoma asso	1.35e+02
38	78	14.0	727	21	W11719	C-Delta-1 polypeptide	1.35e+02
39	78	14.0	740	21	W00876	C-Delta-1 polypeptide	1.35e+02
40	78	14.0	1239	9	R45945	Glutamic acid recepto	1.35e+02
41	77	13.8	528	10	R54206	snbR gene product inv	1.57e+02
42	76	13.6	318	38	W81969	Mouse EI24 protein #2	1.83e+02
43	76	13.6	481	22	W16829	Recombinant endotoxin	1.83e+02
44	76	13.6	481	13	R68922	Lipopolysaccharide bi	1.83e+02
45	76	13.6	1045	39	W86354	Human DNAX toll-like	1.83e+02

ALIGNMENTS

RESULT 1
ID W88498 standard; Protein; 78 AA.
AC W88498;
DT 30-MAR-1999 (first entry)
DE Human stomach carcinoma HP10408-encoded transmembrane protein.
KW Transmembrane protein; HP10408; human; stomach cancer.
OS Homo sapiens.
PN WO9855508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI; 99-045730/04.
DR N-PSDB; V84366.
PT New human proteins containing transmembrane domains and their
PT encoding sequences - useful in the preparation of antibodies and
PT large-scale protein production, gene diagnosis, and gene therapy
PS Claim 1; Page 135; 178pp; English.
CC This is the amino acid sequence of a novel transmembrane protein
CC encoded by human stomach cancer cDNA clone HP10408 (see V84366).
CC The encoded protein has a putative signal sequence and a putative
CC internal transmembrane domain. The invention provides nucleotide
CC sequences (see V84359-76) coding for 18 transmembrane proteins
CC (see W88491-508), vectors containing such polynucleotides, and
CC eukaryotic cells containing the vectors. The proteins can be
CC used as antigens or as compositions in the preparation of
CC antibodies against the proteins. The polynucleotides can be used
CC as probes for gene diagnosis, and as gene sources for gene therapy
CC and large-scale production of proteins encoded by the cDNA. The
CC host cells are used for the detection of ligands corresponding to
CC the expressed proteins, and the screening of low mol.wt. medicines.
SQ Sequence 78 AA;

Query Match 100.0%; Score 558; DB 39; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.84e-38;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mgsglplvllltllgsshgtgpgmtlqlklkesfltnssyessflelleklclllhlpsg 60

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MGSGLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLLLHLPSG 60

Db     61 tsvtlhharsqhhvvcnt 78
          ||||||||||||||||||
Qy     61 TSVTLHHARSQHHVVCNT 78


RESULT    2
ID   W70327 standard; Protein; 46 AA.
AC   W70327;
DT   21-DEC-1998  (first entry)
DE   Secreted protein FB78_1.
KW   Secreted protein; DU123_1; human.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         8..20
FT                   /note= "predicted leader/signal sequence, or
FT                         transmembrane domain"
PN   WO9838209-A2.
PD   03-SEP-1998.
PF   25-FEB-1998; U03697.
PR   24-FEB-1998; US-028724.
PR   26-FEB-1997; US-805819.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 98-481139/41.
DR   N-PSDB; V33199.
PT   New isolated polynucleotide(s) and encoded polypeptide(s) -
PT   obtained from human foetal kidney, adult colon, adult brain, foetal
PT   brain and placenta cDNA libraries.
PS   Claim 36; Page 83; 103pp; English.
CC   This is the amino acid sequence of novel human secreted protein
CC   FB78_1, as deduced from a full-length cDNA clone (see V33199)
CC   obtained from a human adult placenta cDNA library.  Database
CC   searching revealed some similarity between FB78_1 and some known
CC   sequences.  The invention provides new isolated polynucleotides
CC   (see V33190-99), from human foetal kidney, adult colon, adult brain,
CC   foetal brain and placenta cDNA libraries, that code for secreted
CC   proteins (see W70319-27).  The clones can be used for recombinant
CC   production of the polypeptides, which may have activities such as
CC   e.g. nutritional activity, cytokine and cell proliferation or
CC   differentiation activity, immunostimulant or immunosuppressive,
CC   haematopoiesis regulating activity, tissue growth activity, activin
CC   or inhibin activity, chemotactic or chemokinetic activity,
CC   haemostatic and thrombolytic activity, receptor/ligand activity,
CC   antiinflammatory activity, cadherin/tumour invasion suppressor
CC   activity, tumour inhibition activity, or other activities.
SQ   Sequence   46 AA;

  Query Match           15.4%;  Score 86;  DB 36;  Length 46;
  Best Local Similarity  56.5%;  Pred. No. 3.89e+01;
  Matches    13;  Conservative    6;  Mismatches   3;  Indels   1;  Gaps   1;

Db      6 gaalpillllalrgtfhgarpg 28
          |::|| |:|||:| |: ||: ||
Qy      2 GSGLP-LVLLLTLLGSSHGTGPG 23


RESULT    3
ID   W73408 standard; Protein; 47 AA.
AC   W73408;
DT   19-FEB-1999  (first entry)
DE   Human secreted protein encoded by Gene No. 12.
KW   Secreted protein; human; protein therapy; gene therapy; blood disorder;
KW   pathological condition; diagnosis; cancer; neurological disorder;
KW   developmental abnormality; foetal deficiency; leukaemia; hepatic disease;
KW   immune system disorder; Alzheimer's disease; cognitive disorder;
KW   schizophrenia; prostate disease; autoimmune disorder; AIDS.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 47
FT                   /note= "unspecified amino acid"
PN   WO9854206-A1.
PD   03-DEC-1998.
PF   28-MAY-1998; U10868.
PR   29-AUG-1997; US-056296.
PR   30-MAY-1997; US-044039.
PR   30-MAY-1997; US-048093.
PR   30-MAY-1997; US-048101.
PR   30-MAY-1997; US-048190.
PR   30-MAY-1997; US-048356.
PR   30-MAY-1997; US-050935.
PR   29-AUG-1997; US-056250.
PR   29-AUG-1997; US-056293.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Carter KC, Dillon PJ, Endress GA, Feng P, Ni J,
PI   Rosen CA, Ruben SM, Yu G;
DR   WPI; 99-070209/06.
DR   N-PSDB; V08822.
PT   New isolated human genes - useful for diagnosis and treatment of,
PT   e.g. cancers, neurological disorders, immune diseases, developmental
PT   disorders or blood disorders
PS   Claim 11; Page 152-153; 188pp; English.
CC   This sequence is encoded by a cDNA of the invention, designated
CC   Gene No. 12. This sequence represents a human secreted protein, and is
CC   expressed in activated neutorphils, endothelial cells, T-cells and
CC   to a lesser extent in brain and liver.
CC   The DNA sequences of the invention and their corresponding secreted
CC   polypeptides are useful for preventing, treating or ameliorating medical
CC   conditions, e.g. by protein or gene therapy. Also pathological conditions
CC   can be diagnosed by determining the amount of the new polypeptides in a
CC   sample or by determining the presence of mutations in the DNA sequences.
CC   Specific uses are described for each of the DNA sequences and the encoded
CC   proteins, based on which tissues they are most highly expressed in, and
CC   include developing products for the diagnosis or treatment of cancer,
CC   tumours, neurological disorders, developmental abnormalities and foetal
CC   deficiencies, blood disorders, leukaemias, diseases of the immune system
CC   (including allergies or asthma), hepatic disease, Alzheimer's and
CC   cognitive disorders, schizophrenia, prostate diseases, autoimmune
CC   disorders and AIDS. The polypeptides are also useful for identifying
CC   their binding partners.
SQ   Sequence   47 AA;

  Query Match           15.4%;  Score 86;  DB 38;  Length 47;
  Best Local Similarity  56.5%;  Pred. No. 3.89e+01;
  Matches    13;  Conservative    6;  Mismatches   3;  Indels   1;  Gaps   1;

Db      6 gaalpillllalrgtfhgarpg 28
          |::|| |:|||:| |: ||: ||
Qy      2 GSGLP-LVLLLTLLGSSHGTGPG 23


RESULT    4
ID   W02702 standard; peptide; 325 AA.
AC   W02702;
DT   13-NOV-1996  (first entry)
DE   G-protein coupled bovine adrenal angiotensin II type-1 receptor.
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW   muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW   odorant; cytomegalovirus; serotonergic.
OS   Bos taurus.
PN   US5508384-A.
PD   16-APR-1996.
PF   10-SEP-1992; 943236.
PR   10-SEP-1992; US-943236.
PR   09-SEP-1993; US-118270.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
DR   WPI; 96-208785/21.
PT   New dopamine receptor peptide - useful as antipsychotic agent, e.g.
PT   for treating schizophrenia
PS   Disclosure; Column 129-132; 184pp; English.
CC   Proteins W02657-W02730 represent a range of G-protein coupled receptor
```

```
CC   (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other GPR proteins. The receptor proteins
CC   were used to design polypeptides, pref. based on the transmembrane
CC   domains, for use in G-protein coupled receptor ligand binding assays.
CC   The polypeptide fragments retain biological activity such as binding a
CC   GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999
CC   for examples of polypeptide fragments). The polypeptide fragments can
CC   be used in compositions for treating subjects suffering from a pathology
CC   related to a GPR abnormality e.g. a psychotic disorder such as
CC   schizophrenia.
SQ   Sequence   325 AA;

  Query Match           15.2%;  Score 85;  DB 19;  Length 325;
  Best Local Similarity  44.4%;  Pred. No. 4.56e+01;
  Matches    12;  Conservative    5;  Mismatches   8;  Indels   2;  Gaps   2;

Db     27 ymklktyasvfllnlaladlcflltlp 53
          ::   :| | || :| |  ||:|| ||
Qy     34 FLTNSSYESSFL-EL-LEKLCLLLHLP 58


RESULT    5
ID   R48730 standard; Protein; 325 AA.
AC   R48730;
DT   06-JUN-1996  (first entry)
DE   G-protein coupled bovine adrenal angiotensin II type-1 receptor protein.
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   psychotic disorder; schizophrenia; dopamine; cAMP; adenosine; thrombin;
KW   muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
KW   rhodopsin; opsin; odorant; cytomegalovirus.
OS   Bos taurus.
PN   WO9405695-A1.
PD   17-MAR-1994.
PF   09-SEP-1993; U08528.
PR   10-SEP-1992; US-943236.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Murphy RB,  Schuster DI;
DR   WPI; 94-101120/12.
PT   Polypeptides of G-coupled receptor proteins (GPRs) - useful for
PT   binding GPR ligands or modulating GPR binding
PS   Disclosure; Page 106-107; 160pp; English.
CC   Proteins R48685-R48758 represent a range of G-protein coupled receptor
CC   proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other G-protein coupled receptors.  The
CC   receptor proteins were used to design polypeptides, pref. based on the
CC   transmembrane domains, for use in G-protein coupled receptor ligand
CC   binding assays.  The polypeptide fragments retain biological activity
CC   such as binding a GPR ligand or modulating GPR ligand binding to a GPR
CC   (see R48759-R48758, R50569-R50807 and R89189-R89195 for examples of
CC   polypeptide fragments).  The polypeptide fragments can be used in
CC   compositions for treating subjects suffering from a pathology related to
CC   a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
SQ   Sequence   325 AA;

  Query Match           15.2%;  Score 85;  DB 16;  Length 325;
  Best Local Similarity  44.4%;  Pred. No. 4.56e+01;
  Matches    12;  Conservative    5;  Mismatches   8;  Indels   2;  Gaps   2;

Db     27 ymklktyasvfllnlaladlcflltlp 53
          ::   :| | || :| |  ||:|| ||
Qy     34 FLTNSSYESSFL-EL-LEKLCLLLHLP 58


RESULT    6
ID   R47338 standard; Protein; 194 AA.
AC   R47338;
DT   01-JUL-1994  (first entry)
DE   Peptide fragment of tetracycline transporter protein.
KW   Vesicle membrane transport protein; gene therapy; screening;
KW   Parkinsons disease; neurotoxin; identification; detection;
KW   antibody; probe;chromaffin granule amine transporter protein;
KW   resistance; tetracycline.
OS   Escherichia coli.
PN   WO9325699-A.
PD   23-DEC-1993.
PF   11-JUN-1993; U05704.
PR   11-JUN-1992; US-899074.
PR   30-JUL-1992; US-923096.
PA   (REGC ) UNIV CALIFORNIA.
PI   Edwards RH;
DR   WPI; 94-007556/01.
PT   New mammalian vesicle membrane transport protein - and corresp.
PT   DNA, vectors, transformed cells and antibodies, for diagnosis and
PT   treatment of neurological disorders, e.g. Parkinson disease
PS   Example 2; Page 111-112; 181pp; English.
CC   This sequence of the tetracycline transporter protein showed a
CC   definite homology with the chromaffin granule amine transport
CC   protein (CGAT) of rat. The cDNA encoding the chromaffin granule
CC   amine transport protein is useful in gene therapy and as a probe
CC   for detecting genomic sequences.  The transport protein is used for
CC   screening cytotoxic compounds implicated in Parkinsons disease,
CC   diseases associated with activity of neurotoxins or psychiatric
CC   disorders and to identify compounds which selectively inhibit or
CC   activate its action.  Antibodies raised against the transport
CC   protein are useful as immunoassay reagents for detecting the protein
CC   and as affinity reagents for purification.
SQ   Sequence   194 AA;

  Query Match           14.9%;  Score 83;  DB 9;  Length 194;
  Best Local Similarity  31.0%;  Pred. No. 6.23e+01;
  Matches    18;  Conservative   13;  Mismatches  23;  Indels   4;  Gaps   4;

Db      1 mnsstkialvitlldam-giglimpvlptllre-fiasedianhfgvlla-lyalmqv 55
            | |:  : |::||| :  | |  |  |   |:| |:::    : |  ||  |  |:::
Qy      1 MGSGLPLVLLLTLLGSSHGTGPGMT-LQLKLKESFLTNSSYESSFLELLEKLCLLLHL 57


RESULT    7
ID   W57046 standard; Protein; 411 AA.
AC   W57046;
DT   20-AUG-1998  (first entry)
DE   Mouse apoptosis inducing receptor.
KW   Apoptosis inducing receptor; AIR protein; mouse; cell death regulator;
KW   Type I transmembrane protein; tumour cell death; autoimmune disease;
KW   therapy.
OS   Mus sp.
PN   WO9814565-A1.
PD   09-APR-1998.
PF   03-OCT-1997; U17876.
PR   04-OCT-1996; US-044456.
PA   (IMMV ) IMMUNEX CORP.
PI   Perkins PA;
DR   WPI; 98-240077/21.
DR   N-PSDB; V28701.
PT   DNA encoding apoptosis inducing receptor - which is Type I
PT   transmembrane protein, useful for regulating cell death
PS   Claim 16; Page 33-35; 45pp; English.
CC   This sequence is the mouse apoptosis inducing receptor (AIR) of the
CC   invention. AIR is a Type I transmembrane protein, soluble forms of which
CC   can be used to regulate cell death in a therapeutic setting. Soluble AIR
CC   can also be used in vitro to block apoptosis or AIR-expressing cells, or
CC   to screen agonists or antagonists of AIR activity. The cytoplasmic domain
CC   of AIR can be used to develop assays for inhibitors of AIR-induced cell
CC   death, which is useful to regulate cell death in a therapeutic setting as
CC   well as in vitro. Agonists of AIR activity can be used to kill tumour
CC   cells that express AIR, or T cells expressing AIR in autoimmune diseases.
SQ   Sequence   411 AA;

  Query Match           14.9%;  Score 83;  DB 31;  Length 411;
  Best Local Similarity  57.1%;  Pred. No. 6.23e+01;
  Matches    12;  Conservative    5;  Mismatches   3;  Indels   1;  Gaps   1;

Db     16 lplllllllllggqqqg-gms 35
          |||:||| || :::| | ||:
```

```
Qy      5 LPLVLLLTLLGSSHGTGPGMT 25


RESULT    8
ID   R26954 standard; Protein; 110 AA.
AC   R26954;
DT   10-FEB-1993  (first entry)
DE   Human T lymphocyte receptor V-alpha22 segment.
KW   TCR; IGR a 12; variable region; immune system modulation;
KW   T cell subtype.
OS   Homo sapiens.
PN   WO9213949-A.
PD   20-AUG-1992.
PF   07-FEB-1992; F00111.
PR   08-FEB-1991; FR-001487.
PR   12-APR-1991; FR-004527.
PA   (ROUS ) ROUSSEL-UCLAF.
PI   Ferradini L, Hercend T, Roman-roman S, Triebel F;
DR   WPI; 92-300035/36.
DR   N-PSDB; Q28129.
PT   Nucleotide sequences and their monoclonal antibodies and
PT   oligo:nucleotide primers - encode variable alpha-chain regions of
PT   human T-lymphocyte receptors, for studying immune responses and
PT   for therapy
PS   Claim 8; Page 35; 65pp; French.
CC   RNA was isolated from peripheral lymphocytes and converted to cDNA.
CC   The cDNA was amplified by anchored PCR using C-alpha and polyC
CC   primers, then amplified again using a different C-alpha specific
CC   primer. The amplified product was SacII-restricted, inserted into
CC   Bluescript SK+ vector and used to transform E.coli XL7-blue.
CC   Transformants were screened with a C-alpha specific probe and DNA
CC   from positive clones was sequenced in the C-alpha region. The
CC   sequence designated "IGR a 12" comprises the complete coding region
CC   of a gene of the subfamily V alpha 22; this subfamily was
CC   previously identified by the partial sequence (113bp) AC9 (Klein,M.H.,
CC   et al., Proc.Natl.Acad.Sci. USA 84:6884, 1987). The peptide encoded
CC   by it can be used to block specific T cell epitopes or in vaccines.
CC   See Q28120-Q28172.
SQ   Sequence   110 AA;

  Query Match           14.7%;  Score 82;  DB 5;  Length 110;
  Best Local Similarity  34.0%;  Pred. No. 7.28e+01;
  Matches    18;  Conservative   12;  Mismatches  18;  Indels   5;  Gaps   5;

Db      1 mnyspglvslillllgrtrgdsvtqmegpvtlseeafltinctytatgypslf 53
          |  :  || |:| ||| ::| :   |   : | | :|| | :|  :::  |:
Qy      1 MGSGLPLV-LLLTLLGSSHGTG-PGMTLQLKLKE-SFLT-NSSY-ESSFLELL 48


RESULT    9
ID   W21674 standard; Protein; 169 AA.
AC   W21674;
DT   29-SEP-1997  (first entry)
DE   Human mitochondrial electron transport chain subunit CII-3.
KW   Mammalian artificial chromosome; MAC; selectable marker; CII-3;
KW   mitochondrial electron transport chain complex II.
OS   Homo sapiens.
PN   WO9716533-A1.
PD   09-MAY-1997.
PF   29-OCT-1996; U17476.
PR   31-OCT-1995; US-550717.
PA   (REGC ) UNIV CALIFORNIA.
PI   Scheffler IE;
DR   WPI; 97-272103/24.
DR   N-PSDB; T72466.
PT   New mammalian artificial chromosomes - comprising a mammalian
PT   centromere and a unique cloning site, used for stable expression of
PT   large fragments of DNA
PS   Disclosure; Page 54-55; 71pp; English.
CC   CII-3 (W21674) comprises a subunit of complex II of the human
CC   mitochondrial electron transport chain.  It is the expression
CC   product of the CII-3 gene identified in mammalian artificial
CC   chromosome MAC-8.2.3, which is contained in the human-hamster
CC   hybrid cell line XEW8.2.3 (ATCC CRL 11991).  A portion of the
CC   CII-3 gene (see also T72461-65) or of CII-3 cDNA (T72466) can be
CC   utilised as a unique cloning site and selectable marker in an MAC,
CC   allowing the site-specific integration of an exogenous nucleic acid
CC   sequence into the MAC.  The MAC can be used for stable expression
CC   of large fragments of DNA and also for the production of transgenic
CC   animals.
SQ   Sequence   169 AA;

  Query Match           14.7%;  Score 82;  DB 23;  Length 169;
  Best Local Similarity  28.0%;  Pred. No. 7.28e+01;
  Matches    14;  Conservative   20;  Mismatches  13;  Indels   3;  Gaps   3;

Db     62 slpmamsichrgtgialsagvsl-fgms-alllpgnfes-ylelvkslcl 108
          :||: : :   |:: : ::|::| : :  ::|  :::|| :|||:  |||
Qy      4 GLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCL 53


RESULT   10
ID   W59924 standard; Protein; 401 AA.
AC   W59924;
DT   07-DEC-1998  (first entry)
DE   Human 7-transmembrane receptor HNFDY20.
KW   HNFDY20; G-protein coupled receptor; human; infection; HIV; pain;
KW   cancer; anorexia; asthma; Parkinson's disease; acute heart failure;
KW   hypotension; hypertension; urinary retention; osteoporosis;
KW   angina pectoris; myocardial infarction; ulcer; allergy;
KW   benign prostatic hypertrophy; psychosis; anxiety; schizophrenia;
KW   manic depression; delirium; dementia; mental retardation;
KW   dyskinesia; Huntingdon's disease; Gilles de la Tourette's syndrome;
KW   therapy; diagnosis.
OS   Homo sapiens.
PN   EP-866126-A1.
PD   23-SEP-1998.
PF   16-FEB-1998; 301122.
PR   19-MAR-1997; US-820521.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Bergsma DJ, Fuetterer WS, Mao JY, Sathe GM;
DR   WPI; 98-482962/42.
DR   N-PSDB; V53631.
PT   New polynucleotides and polypeptides encoding a novel human
PT   7-transmembrane receptor - useful for diagnosing and treating e.g.
PT   cancer, osteoporosis and Parkinson's disease and infections caused
PT   by HIV-1 or -2.
PS   Claim 1; Page 18-19; 24pp; English.
CC   This polypeptide comprises HNFDY20, a novel human 7-transmembrane
CC   G-protein coupled receptor that shows about 30.8% identity in 299
CC   amino acid residues with the thrombin receptor.  Its amino acid
CC   sequence was deduced from an isolated HNFDY20 polynucleotide
CC   sequence (see V53631).  The invention relates to HNFDY20
CC   polypeptides and recombinant materials and methods for their
CC   production.  It also provides methods for using such polypeptides
CC   and HNFDY20 polynucleotides for treatment of infections such as
CC   bacterial, fungal, protozoan and particularly HIV-1 or HIV-2
CC   infections, and conditions including pain, cancers, anorexia,
CC   asthma, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   myocardial infarction, ulcers, allergies, benign prostatic
CC   hypertrophy, and psychotic and neurological disorders, including
CC   anxiety, schizophrenia, manic depression, delirium, dementia,
CC   severe mental retardation and dyskinesias, such as Huntingdon's
CC   disease or Gilles de la Tourette's syndrome.  Gene therapy using
CC   RNA encoding HNFDY20 can be used to treat conditions caused by
CC   under-expression of the protein.  The invention also relates to
CC   methods of identifying agonists and antagonists and for using
CC   such compounds to treat conditions associated with HNFDY20
CC   imbalance.  Diagnostic assays for detecting diseases associated
CC   with inappropriate HNFDY20 activity or levels are also provided.
SQ   Sequence   401 AA;

  Query Match           14.5%;  Score 81;  DB 35;  Length 401;
  Best Local Similarity  38.1%;  Pred. No. 8.50e+01;
  Matches     8;  Conservative    9;  Mismatches   4;  Indels   0;  Gaps   0;
```

```
Db    332 yfsssgfqadfhellrrlcgl 352
          ::::|:::: | ||| :|| |
Qy     34 FLTNSSYESSFLELLEKLCLL 54
```

RESULT 11
ID W10038 standard; Protein; 1253 AA.
AC W10038;
DT 07-FEB-1998 (first entry)
DE Mad binding protein, mSinA.
KW murine; mSinA; mammalian homologue; Saccharomyces cerevisiae; repressor;
KW Sin3; Mad; Max; mSin:Mad complex; mSin:Mad:Max complex; Myc; promoter;
KW basic helix-loop-helix zipper protein; compete; DNA-binding;
KW Myc:Max complex; activate; transcription; gene regulation.
OS Mus musculus.
FH Key Location/Qualifiers
FT Misc_difference 10
FT /label= unknown
FT /note= "encoded by TAG"
FT Misc_difference 1238
FT /label= unknown
FT /note= "encoded by TAG"
PN US5624818-A.
PD 29-APR-1997.
PF 01-JUN-1994; 252966.
PR 01-JUN-1994; US-252966.
PR 19-SEP-1991; US-756195.
PR 23-JUN-1992; US-903710.
PR 01-APR-1994; US-222638.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Ayer DE, Eisenman RN;
DR WPI; 97-258216/23.
DR N-PSDB; T70126.
PT mSin nucleic acids encoding recombinant polypeptide(s) that
PT associate with Mad polypeptide - are possible homologues of S.
PT cerevisiae general repressor protein Sin3
PS Example 12; Fig 23A-C; 111pp; English.
CC This sequence represents the murine protein, designated mSinA, which may
CC be a mammalian homologue of the Saccharomyces cerevisiae general
CC repressor protein Sin3. The mSin protein associates with a Mad
CC polypeptide to form a mSin:Mad complex, which preferably associates
CC with a Max polypeptide to form a mSin:Mad:Max complex which binds to a
CC nucleotide sequence comprising CACGTG. .Mad is a basic helix-loop-helix
CC (bHLH) zipper protein which can compete with Myc by forming sequence-
CC specific DNA-binding heterocomplexes with Max. Mad:Max complexes repress,
CC while Myc:Max complexes activate, transcription from promoters containing
CC proximal CACGTG binding sites for these proteins. Expression of Mad is
CC closely linked to differentation in at least two distinct cell lineages.
CC The switch from Myc:Max to Mad:Max complexes may reflect the repression
CC of transcription of Myc regulated genes by Mad. The DNA, vectors and host
CC cells of the invention are useful for the recombinant production of mSin
CC proteins useful in elucidation of Mad repressor functions.
SQ Sequence 1253 AA;

Query Match 14.5%; Score 81; DB 25; Length 1253;
Best Local Similarity 27.5%; Pred. No. 8.50e+01;
Matches 11; Conservative 14; Mismatches 14; Indels 1; Gaps 1;

```
Db    949 vlgikrdksdspaiqlrlkepmdvdvedyypafldmvrsl 988
          :||  :  : : ::||:|||:  :    | ::||:::  |
Qy     13 LLGSSHGTGPGMTLQLKLKESF-LTNSSYESSFLELLEKL 51
```

RESULT 12
ID W10040 standard; Protein; 1261 AA.
AC W10040;
DT 07-FEB-1998 (first entry)
DE Mad binding protein, mSinA9.
KW murine; mSinA; mammalian homologue; Saccharomyces cerevisiae; repressor;
KW Sin3; Mad; Max; mSin:Mad complex; mSin:Mad:Max complex; Myc; promoter;
KW basic helix-loop-helix zipper protein; compete; DNA-binding;
KW Myc:Max complex; activate; transcription; gene regulation.
OS Mus musculus.
FH Key Location/Qualifiers
FT Misc_difference 10
FT /label= unknown
FT /note= "encoded by TAG"
FT Misc_difference 1247
FT /label= unknown
FT /note= "encoded by TAG"
PN US5624818-A.
PD 29-APR-1997.
PF 01-JUN-1994; 252966.
PR 01-JUN-1994; US-252966.
PR 19-SEP-1991; US-756195.
PR 23-JUN-1992; US-903710.
PR 01-APR-1994; US-222638.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PI Ayer DE, Eisenman RN;
DR WPI; 97-258216/23.
DR N-PSDB; T70128.
PT mSin nucleic acids encoding recombinant polypeptide(s) that
PT associate with Mad polypeptide - are possible homologues of S.
PT cerevisiae general repressor protein Sin3
PS Example 12; Fig 29A-C; 111pp; English.
CC This sequence represents the murine protein, designated mSinA9, which
CC may be a mammalian homologue of the Saccharomyces cerevisiae general
CC repressor protein Sin3. The mSin protein associates with a Mad
CC polypeptide to form a mSin:Mad complex, which preferably associates
CC with a Max polypeptide to form a mSin:Mad:Max complex which binds to a
CC nucleotide sequence comprising CACGTG. Mad is a basic helix-loop-helix
CC (bHLH) zipper protein which can compete with Myc by forming sequence-
CC specific DNA-binding heterocomplexes with Max. Mad:Max complexes repress,
CC while Myc:Max complexes activate, transcription from promoters containing
CC proximal CACGTG binding sites for these proteins. Expression of Mad is
CC closely linked to differentation in at least two distinct cell lineages.
CC The switch from Myc:Max to Mad:Max complexes may reflect the repression
CC of transcription of Myc regulated genes by Mad. The DNA, vectors and host
CC cells of the invention are useful for the recombinant production of mSin
CC proteins useful in elucidation of Mad repressor functions.
SQ Sequence 1261 AA;

Query Match 14.5%; Score 81; DB 25; Length 1261;
Best Local Similarity 27.5%; Pred. No. 8.50e+01;
Matches 11; Conservative 14; Mismatches 14; Indels 1; Gaps 1;

```
Db    949 vlgikrdksdspaiqlrlkepmdvdvedyypafldmvrsl 988
          :||  :  : : ::||:|||:  :    | ::||:::  |
Qy     13 LLGSSHGTGPGMTLQLKLKESF-LTNSSYESSFLELLEKL 51
```

RESULT 13
ID R15138 standard; Protein; 398 AA.
AC R15138;
DT 17-FEB-1992 (first entry)
DE Human serotonin 1D receptor encoded by gene 11.
KW 5-HT (1D); Parkinson's Disease; migraine; anxiety; eating disorder;
KW G-protein; 5-hydroxytryptamine.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 1..49
FT /label= extracellular
FT domain 50..73
FT /label= transmembrane-1
FT domain 74..86
FT /label= cytoplasmic
FT domain 87..110
FT /label= transmembrane-2
FT domain 111..121
FT /label= extracellular
FT domain 122..145
FT /label= transmembrane-3
FT domain 146..164
FT /label= cytoplasmic
FT domain 165..188

```
FT                  /label= transmembrane-4
FT   domain         189..205
FT                  /label= extracellular
FT   domain         206..229
FT                  /label= transmembrane-5
FT   domain         230..315
FT                  /label= cytoplasmic_loop
FT   domain         316..339
FT                  /label= transmembrane-6
FT   domain         340..348
FT                  /label= extracellular
FT   domain         349..372
FT                  /label= transmembrane-7
FT   region         373..398
FT                  /label= cytoplasmic_tail
PN   WO9117174-A.
PD   14-NOV-1991.
PF   08-MAY-1991; U03200.
PR   08-MAY-1990; US-520716.
PA   (NEUR-) NEUROGENETIC CORP.
PI   Weinshank RL, Branchek T, Hartig PR;
DR   WPI; 91-353715/48.
DR   N-PSDB; Q14836.
PT   Nucleic acid encoding 5HT-1D receptors and their antibodies -
PT   used to treat and diagnose conditions caused by abnormal 5HT-1D
PT   receptor expression e.g. dementia
PS   Claim 7; Fig 4; 90pp; English.
CC   There are two subtypes of 5HT(1D) receptors encoded by gene 8-30-84
CC   and gene 11, respectively. A full-length clone corresponding to
CC   gene 11 was isolated from a human placental library. The
CC   amino acid sequence deduced from the nucleotide sequence predicts a
CC   mol.wt. 44,333 protein. Comparison of this protein sequence with
CC   previously characterised neurotransmitter receptors suggests that
CC   it is a new member of the G protein-coupled receptor family.
CC   See Q14836 for gene 11.
SQ   Sequence   398 AA;

  Query Match           14.3%;  Score 80;  DB 3;  Length 398;
  Best Local Similarity  23.6%;  Pred. No. 9.93e+01;
  Matches    17;  Conservative   22;  Mismatches  32;  Indels   1;  Gaps   1;

Db     52 lvmllalitlattlsnafviatvyrtrklhtpanyliasldvtdllvsilvipistmytv 111
          ||:||:|:  :   : :  :    :   | | ::|  : |:: : |  :| :| :|  |:
Qy      7 LVLLLTLLGSSHGTGPGMTLQLKLKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTL 65

Db    112 tdrwtlsqvvcd 123
            ::  :|||:
Qy     66 HHARSQHHVVCN 77


RESULT   14
ID   R10052 standard; Protein; 713 AA.
AC   R10052;
DT   19-MAR-1991  (first entry)
DE   Cyclomaltodextrin glucotransferase 17-1.
KW   CGTase.
OS   Bacillus strain 17-1 (ATCC 31007)
FH   Key             Location/Qualifiers
FT   peptide         1..27
FT                   /label= signal peptide
FT   peptide         28..44
FT                   /note= "confirmed by N-terminal Edman degradation"
FT   region          1..260
FT                   /label= U1
FT   region          261..439
FT                   /label= U2
FT   region          440..536
FT                   /label= U3
FT   region          537..713
FT                   /label= U4
PN   J02286080-A.
PD   26-NOV-1990.
PF   28-APR-1989; 110603.
PR   28-APR-1989; JP-110603.
PA   (RIKA ) RIKAGAKU KENKYUSHO.
DR   WPI; 91-012216/02.
DR   Q-PSDB; Q10164.
PT   New cyclo:malto:dextrin glucanotransferase - has DNA sequence
PT   coding amino acid sequence.
PS   Disclosure; Fig 6; 16pp; Japanese.
CC   The sequence is deduced from a DNA from pUCP1.  The sequence is
CC   divided into four units (U1-U4) by restriction sites in the DNA.
CC   By combining these with similar units (S1-S4) from CGTase 38-2
CC   (R10051), e.g.S1-U2-U3-S4 or U1-S2-U3-S4, enzymes with properties
CC   of both strains can be produced.   The hybrids compensate for
CC   defects in the individual enzymes and can be designed to display
CC   optimum characteristics of pH and temperature stability.
CC   See also R10052.
SQ   Sequence   713 AA;

  Query Match           14.3%;  Score 80;  DB 2;  Length 713;
  Best Local Similarity  34.2%;  Pred. No. 9.93e+01;
  Matches    13;  Conservative   14;  Mismatches  8;  Indels   3;  Gaps   3;

Db     12 alslslalsllgpahaa-pdtsvs-n-kqnfstdviyq 46
          :|:| | |:|||::|:: |  ::  : |::| |:  |:
Qy      4 GLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYE 41


RESULT   15
ID   R06110 standard; protein; 713 AA.
AC   R06110;
DT   20-DEC-1990  (first entry)
DE   Sequence of cyclomaltodextrin glucanotransferase (CGTase)
DE   having alkaline optimum pH
KW   Enzyme.
PN   J02000455-A.
PD   05-JAN-1990.
PF   28-MAR-1988; 105945.
PR   10-JUL-1987; JP-172722.
PR   8-MAR-1988; JP-105945.
PA   (RIKA) Rikagaku Kenkyusho.
DR   WPI; 90-047991/07.
DR   N-PSDB; Q01810.
PT   Cyclo:malto:dextrin glucano:transferase coding DNA sequence -
PT   including corresp. plasmid and microbe transformed by plasmid
PS   Disclosure; Page 831; 16pp; Japanese.
CC   DNA, the plasmid contg. it, and any microorganism transformed with the
CC   plasmid are new. Transformed strains show CGTase activity after 16 hrs.
CC   culture on LB plate. 38-2 CGTase and 17-1 CGTase genes are each stably
CC   retained in E.coli cells translated to CGTase efficiently. Large amts. of
CC   polypeptide with CGTase activity can be obtd.
SQ   Sequence   713 AA;

  Query Match           14.3%;  Score 80;  DB 1;  Length 713;
  Best Local Similarity  34.2%;  Pred. No. 9.93e+01;
  Matches    13;  Conservative   14;  Mismatches  8;  Indels   3;  Gaps   3;

Db     12 alslslalsllgpahaa-pdtsvs-n-kqnfstdviyq 46
          :|:| | |:|||::|:: |  ::  : |::| |:  |:
Qy      4 GLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYE 41


Search completed: Fri Oct 22 18:42:22 1999
Job time : 55 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 18:45:15 1999; MasPar time 2.38 Seconds
383.184 Million cell updates/sec
Tabular output not generated.

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect Score: 558
Sequence: 1 MGSGLPLVLLLTLLGSSHGT.....SGTSVTLHHARSQHHVVCNT 78

Scoring table: PAM 150
Gap 11

Searched: 119857 seqs, 11713122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 25.509; Variance 122.762; scale 0.208

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	85	15.2	325	1	US-08-118-	Sequence 51, Applicati	2.00e+01
2	85	15.2	325	3	PCT-US93-0	Sequence 51, Applicati	2.00e+01
3	83	14.9	194	1	US-08-063-	Sequence 8, Applicatio	2.75e+01
4	83	14.9	194	3	PCT-US93-0	Sequence 8, Applicatio	2.75e+01
5	82	14.7	169	1	US-08-741-	Sequence 2, Applicatio	3.23e+01
6	81	14.5	401	2	US-08-820-	Sequence 2, Applicatio	3.78e+01
7	81	14.5	1253	1	US-08-252-	Sequence 12, Applicati	3.78e+01
8	81	14.5	1261	1	US-08-252-	Sequence 18, Applicati	3.78e+01
9	80	14.3	169	2	US-08-828-	Sequence 4, Applicatio	4.43e+01
10	80	14.3	390	1	US-07-817-	Sequence 6, Applicatio	4.43e+01
11	80	14.3	390	1	US-08-117-	Sequence 6, Applicatio	4.43e+01
12	80	14.3	390	3	PCT-US93-0	Sequence 6, Applicatio	4.43e+01
13	80	14.3	390	1	US-08-216-	Sequence 6, Applicatio	4.43e+01
14	80	14.3	398	2	US-08-542-	Sequence 6, Applicatio	4.43e+01
15	80	14.3	398	1	US-08-370-	Sequence 6, Applicatio	4.43e+01
16	80	14.3	1214	2	US-08-231-	Sequence 54, Applicati	4.43e+01
17	80	14.3	1219	2	US-08-231-	Sequence 50, Applicati	4.43e+01
18	80	14.3	1231	2	US-08-231-	Sequence 48, Applicati	4.43e+01
19	80	14.3	1236	2	US-08-231-	Sequence 6, Applicatio	4.43e+01
20	80	14.3	1239	2	US-08-231-	Sequence 52, Applicati	4.43e+01
21	80	14.3	1244	2	US-08-231-	Sequence 46, Applicati	4.43e+01
22	79	14.2	429	1	US-08-339-	Sequence 33, Applicati	5.18e+01
23	78	14.0	120	2	US-08-420-	Sequence 29, Applicati	6.06e+01
24	78	14.0	120	3	PCT-US95-1	Sequence 29, Applicati	6.06e+01
25	78	14.0	1239	1	US-08-026-	Sequence 3, Applicatio	6.06e+01
26	76	13.6	197	3	PCT-US95-0	Sequence 2, Applicatio	8.28e+01
27	76	13.6	197	2	US-08-215-	Sequence 2, Applicatio	8.28e+01
28	76	13.6	317	2	US-09-213-	Sequence 3, Applicatio	8.28e+01
29	76	13.6	318	2	US-08-619-	Sequence 9, Applicatio	8.28e+01
30	76	13.6	340	2	US-09-213-	Sequence 1, Applicatio	8.28e+01
31	76	13.6	477	4	5245013-3	Patent No. 5245013.	8.28e+01
32	76	13.6	477	1	US-07-847-	Sequence 2, Applicatio	8.28e+01
33	76	13.6	477	2	US-08-990-	Sequence 2, Applicatio	8.28e+01
34	76	13.6	481	1	US-08-306-	Sequence 98, Applicati	8.28e+01
35	76	13.6	481	1	US-08-261-	Sequence 4, Applicatio	8.28e+01
36	76	13.6	481	2	US-08-274-	Sequence 6, Applicatio	8.28e+01
37	76	13.6	481	1	US-08-186-	Sequence 2, Applicatio	8.28e+01
38	76	13.6	481	2	US-08-485-	Sequence 98, Applicati	8.28e+01
39	76	13.6	481	3	PCT-US95-0	Sequence 9, Applicatio	8.28e+01
40	76	13.6	481	3	PCT-US95-0	Sequence 98, Applicati	8.28e+01
41	76	13.6	481	3	PCT-US95-0	Sequence 98, Applicati	8.28e+01
42	76	13.6	481	3	PCT-US94-0	Sequence 98, Applicati	8.28e+01
43	76	13.6	481	2	US-08-215-	Sequence 9, Applicatio	8.28e+01
44	76	13.6	481	3	PCT-US94-0	Sequence 4, Applicatio	8.28e+01
45	76	13.6	528	2	US-08-403-	Sequence 21, Applicati	8.28e+01

ALIGNMENTS

```
RESULT    1
ID   US-08-118-270-51     STANDARD;     PRT;    325 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 51, Application US/08118270
XX
CC    Sequence 51, Application US/08118270
CC    Patent No. 5508384
CC     GENERAL INFORMATION:
CC       APPLICANT:  Murphy, Randall B.
CC       APPLICANT:  Schuster, David I.
CC       TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC       TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC       NUMBER OF SEQUENCES:  348
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  BROWDY AND NEIMARK
CC         STREET:  419 Seventh Street, N.W., Suite 300
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/118,270
CC         FILING DATE:  09-SEP-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/943,236
CC         FILING DATE:  10-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Townsend, Kevin G.
CC         REGISTRATION NUMBER:  34,033
CC         REFERENCE/DOCKET NUMBER:  MURPHY=2A
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC         TELEX:  248633
CC     INFORMATION FOR SEQ ID NO:  51:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  325 amino acids
CC         TYPE:  amino acid
```

```
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   325 AA; 37288 MW; 592694 CN;

  Query Match           15.2%;  Score 85;  DB 1;  Length 325;
  Best Local Similarity  44.4%;  Pred. No. 2.00e+01;
  Matches    12;  Conservative    5;  Mismatches   8;  Indels   2;  Gaps   2;

Db     27 YMKLKTYASVFLLNLALADLCFLLTLP 53
          ::   :| | || :| |  ||:|| ||
Qy     34 FLTNSSYESSFL-EL-LEKLCLLLHLP 58


RESULT    2
ID   PCT-US93-08528-51     STANDARD;     PRT;   325 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 51, Application PC/TUS9308528
XX
CC    Sequence 51, Application PC/TUS9308528
CC     GENERAL INFORMATION:
CC       APPLICANT:  New York University
CC       TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
CC       TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
CC       NUMBER OF SEQUENCES:  348
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  BROWDY AND NEIMARK
CC         STREET:  419 Seventh Street, N.W., Suite 300
CC         CITY:  Washington
CC         STATE:  D.C.
CC         COUNTRY:  USA
CC         ZIP:  20004
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/08528
CC         FILING DATE:  09-SEP-1993
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 07/943,236
CC         FILING DATE:  10-SEP-1992
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Townsend, Kevin G.
CC         REGISTRATION NUMBER:  34,033
CC         REFERENCE/DOCKET NUMBER:  MURPHY=2 PCT
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  202-628-5197
CC         TELEFAX:  202-737-3528
CC         TELEX:  248633
CC     INFORMATION FOR SEQ ID NO:  51:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  325 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
SQ   SEQUENCE   325 AA; 37288 MW; 592694 CN;

  Query Match           15.2%;  Score 85;  DB 3;  Length 325;
  Best Local Similarity  44.4%;  Pred. No. 2.00e+01;
  Matches    12;  Conservative    5;  Mismatches   8;  Indels   2;  Gaps   2;

Db     27 YMKLKTYASVFLLNLALADLCFLLTLP 53
          ::   :| | || :| |  ||:|| ||
Qy     34 FLTNSSYESSFL-EL-LEKLCLLLHLP 58


RESULT    3
ID   US-08-063-552-8     STANDARD;     PRT;   194 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application US/08063552
XX
CC    Sequence 8, Application US/08063552
CC    Patent No. 5688936
CC     GENERAL INFORMATION:
CC       APPLICANT:  Edwards, Robert H
CC       TITLE OF INVENTION:  Vesicle Membrane Transport Proteins
CC       NUMBER OF SEQUENCES:  17
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Sheldon & Mak
CC         STREET:  225 South Lake Avenue, Ninth Floor
CC         CITY:  Pasadena
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  91101
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/063,552
CC         FILING DATE:  19930514
CC         CLASSIFICATION:  530
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Farber, Michael B
CC         REGISTRATION NUMBER:  32,612
CC         REFERENCE/DOCKET NUMBER:  9067-1
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (818) 796-4000
CC         TELEFAX:  (818) 795-6321
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  194 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  NO
CC       FRAGMENT TYPE:  internal
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Transposon 10
SQ   SEQUENCE   194 AA; 20868 MW; 201442 CN;

  Query Match           14.9%;  Score 83;  DB 1;  Length 194;
  Best Local Similarity  31.0%;  Pred. No. 2.75e+01;
  Matches    18;  Conservative   13;  Mismatches  23;  Indels   4;  Gaps   4;

Db      1 MNSSTKIALVITLLDAM-GIGLIMPVLPTLLRE-FIASEDIANHFGVLLA-LYALMQV 55
          | |:  : |::||| :  | |  |  |   |:| |:::    : |  ||  |  |:::
Qy      1 MGSGLPLVLLLTLLGSSHGTGPGMT-LQLKLKESFLTNSSYESSFLELLEKLCLLLHL 57


RESULT    4
ID   PCT-US93-05704-8     STANDARD;     PRT;   194 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 8, Application PC/TUS9305704
XX
CC    Sequence 8, Application PC/TUS9305704
CC     GENERAL INFORMATION:
CC       APPLICANT:  Edwards, Robert H
CC       TITLE OF INVENTION:  Vesicle Membrane Transport Proteins
CC       NUMBER OF SEQUENCES:  17
```

```
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Sheldon & Mak
CC         STREET:  225 South Lake Avenue, Ninth Floor
CC         CITY:  Pasadena
CC         STATE:  California
CC         COUNTRY:  USA
CC         ZIP:  91001
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/05704
CC         FILING DATE:  19930611
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Farber, Michael B
CC         REGISTRATION NUMBER:  32,612
CC         REFERENCE/DOCKET NUMBER:  9067-1PCT
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (818) 796-4000
CC         TELEFAX:  (818) 795-6321
CC     INFORMATION FOR SEQ ID NO:  8:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  194 amino acids
CC         TYPE:  AMINO ACID
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  peptide
CC       HYPOTHETICAL:  NO
CC       FRAGMENT TYPE:  internal
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Transposon 10
SQ   SEQUENCE   194 AA; 20868 MW; 201442 CN;

  Query Match           14.9%;  Score 83;  DB 3;  Length 194;
  Best Local Similarity  31.0%;  Pred. No. 2.75e+01;
  Matches    18;  Conservative   13;  Mismatches  23;  Indels   4;  Gaps   4;

Db       1 MNSSTKIALVITLLDAM-GIGLIMPVLPTLLRE-FIASEDIANHFGVLLA-LYALMQV 55
           | |:  : |::||| :  | |  |  |   |:| |:::    : |  ||  |  |:::
Qy       1 MGSGLPLVLLLTLLGSSHGTGPGMT-LQLKLKESFLTNSSYESSFLELLEKLCLLLHL 57


RESULT    5
ID   US-08-741-406-2     STANDARD;     PRT;   169 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08741406
XX
CC    Sequence 2, Application US/08741406
CC    Patent No. 5721118
CC     GENERAL INFORMATION:
CC       APPLICANT:  Scheffler, Immo E.
CC       TITLE OF INVENTION:  Mammalian Artificial Chromosomes and
CC       TITLE OF INVENTION:  Methods of Using Same
CC       NUMBER OF SEQUENCES:  16
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Campbell & Flores LLP
CC         STREET:  4370 La Jolla Village Drive, Suite 700
CC         CITY:  San Diego
CC         STATE:  California
CC         COUNTRY:  United States
CC         ZIP:  92122
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/741,406
CC         FILING DATE:
CC         CLASSIFICATION:  514
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US 08/550,717
CC         FILING DATE:  31-OCT-1995
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Campbell, Cathryn A.
CC         REGISTRATION NUMBER:  31,815
CC         REFERENCE/DOCKET NUMBER:  P-UD 2317
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (619) 535-9001
CC         TELEFAX:  (619) 535-8949
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  169 amino acids
CC         TYPE:  amino acid
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   169 AA; 18610 MW; 162524 CN;

  Query Match           14.7%;  Score 82;  DB 1;  Length 169;
  Best Local Similarity  28.0%;  Pred. No. 3.23e+01;
  Matches    14;  Conservative   20;  Mismatches  13;  Indels   3;  Gaps   3;

Db      62 SLPMAMSICHRGTGIALSAGVSL-FGMS-ALLLPGNFES-YLELVKSLCL 108
           :||: : :   |:: : ::|::| : :  ::|  :::|| :|||:  |||
Qy       4 GLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCL 53


RESULT    6
ID   US-08-820-521-2     STANDARD;     PRT;   401 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 2, Application US/08820521
XX
CC    Sequence 2, Application US/08820521
CC    Patent No. 5942416
CC     GENERAL INFORMATION:
CC       APPLICANT:  Bergsma, Derk
CC       APPLICANT:  Ganesh, Sathe
CC       APPLICANT:  Fuetterer, Wendy
CC       APPLICANT:  Mao, Joyce
CC       TITLE OF INVENTION:  cDNA CLONE HNFDY20 THAT ENCODES
CC       TITLE OF INVENTION:  A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR
CC       NUMBER OF SEQUENCES:  2
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  SmithKline Beecham Corporation
CC         STREET:  709 Swedeland Road
CC         CITY:  King of Prussia
CC         STATE:  PA
CC         COUNTRY:  USA
CC         ZIP:  19406
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/820,521
CC         FILING DATE:  19-MAR-1997
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Han, William T
CC         REGISTRATION NUMBER:  34,344
CC         REFERENCE/DOCKET NUMBER:  GH50011
CC       TELECOMMUNICATION INFORMATION:
```

```
CC         TELEPHONE:  610-270-5219
CC         TELEFAX:  610-270-4026
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  2:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  401 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
SQ   SEQUENCE   401 AA; 44386 MW; 855771 CN;

  Query Match           14.5%;  Score 81;  DB 2;  Length 401;
  Best Local Similarity  38.1%;  Pred. No. 3.78e+01;
  Matches     8;  Conservative    9;  Mismatches   4;  Indels   0;  Gaps   0;

Db    332 YFSSSGFQADFHELLRRLCGL 352
          ::::|:::: | ||| :|| |
Qy     34 FLTNSSYESSFLELLEKLCLL 54


RESULT    7
ID   US-08-252-966B-12     STANDARD;     PRT;   1253 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 12, Application US/08252966B
XX
CC    Sequence 12, Application US/08252966B
CC    Patent No. 5624818
CC     GENERAL INFORMATION:
CC       APPLICANT:  Eisenman, Robert N.
CC       APPLICANT:  Hurlin, Peter J.
CC       APPLICANT:  Ayer, Donald E.
CC       TITLE OF INVENTION:  Regulatory Proteins that Dimerize with
CC       TITLE OF INVENTION:  Mad or Max
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Christensen, O'Connor, Johnson, and KindnessPLLC
CC         STREET:  1420 Fifth Ave., Suite 2800
CC         CITY:  Seattle
CC         STATE:  Washington
CC         COUNTRY:  USA
CC         ZIP:  98101-2347
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/252,966B
CC         FILING DATE:  01-JUN-1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Shelton, Dennis K.
CC         REGISTRATION NUMBER:  26,997.
CC         REFERENCE/DOCKET NUMBER:  FHCR17694
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (206) 682-8100
CC         TELEFAX:  (206) 224-0779
CC     INFORMATION FOR SEQ ID NO:  12:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1253 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC         DESCRIPTION:  translation of msina cDNA; see Figure 23
CC       HYPOTHETICAL:  YES
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Mus musculus
SQ   SEQUENCE   1253 AA; 142589 MW; 7863283 CN;

  Query Match           14.5%;  Score 81;  DB 1;  Length 1253;
  Best Local Similarity  27.5%;  Pred. No. 3.78e+01;
  Matches    11;  Conservative   14;  Mismatches  14;  Indels   1;  Gaps   1;

Db    949 VLGIKRDKSDSPAIQLRLKEPMDVDVEDYYPAFLDMVRSL 988
          :||  :  : : ::||:|||:  :    | ::||:::  |
Qy     13 LLGSSHGTGPGMTLQLKLKESF-LTNSSYESSFLELLEKL 51


RESULT    8
ID   US-08-252-966B-18     STANDARD;     PRT;   1261 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 18, Application US/08252966B
XX
CC    Sequence 18, Application US/08252966B
CC    Patent No. 5624818
CC     GENERAL INFORMATION:
CC       APPLICANT:  Eisenman, Robert N.
CC       APPLICANT:  Hurlin, Peter J.
CC       APPLICANT:  Ayer, Donald E.
CC       TITLE OF INVENTION:  Regulatory Proteins that Dimerize with
CC       TITLE OF INVENTION:  Mad or Max
CC       NUMBER OF SEQUENCES:  19
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Christensen, O'Connor, Johnson, and KindnessPLLC
CC         STREET:  1420 Fifth Ave., Suite 2800
CC         CITY:  Seattle
CC         STATE:  Washington
CC         COUNTRY:  USA
CC         ZIP:  98101-2347
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/252,966B
CC         FILING DATE:  01-JUN-1994
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Shelton, Dennis K.
CC         REGISTRATION NUMBER:  26,997.
CC         REFERENCE/DOCKET NUMBER:  FHCR17694
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  (206) 682-8100
CC         TELEFAX:  (206) 224-0779
CC     INFORMATION FOR SEQ ID NO:  18:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  1261 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC         DESCRIPTION:  translation of msina9 cDNA; see Figure 29A, B, C, D
CC       HYPOTHETICAL:  YES
CC       ORIGINAL SOURCE:
CC         ORGANISM:  Mus musculus
SQ   SEQUENCE   1261 AA; 143711 MW; 7937040 CN;

  Query Match           14.5%;  Score 81;  DB 1;  Length 1261;
  Best Local Similarity  27.5%;  Pred. No. 3.78e+01;
  Matches    11;  Conservative   14;  Mismatches  14;  Indels   1;  Gaps   1;

Db    949 VLGIKRDKSDSPAIQLRLKEPMDVDVEDYYPAFLDMVRSL 988
          :||  :  : : ::||:|||:  :    | ::||:::  |
Qy     13 LLGSSHGTGPGMTLQLKLKESF-LTNSSYESSFLELLEKL 51
```

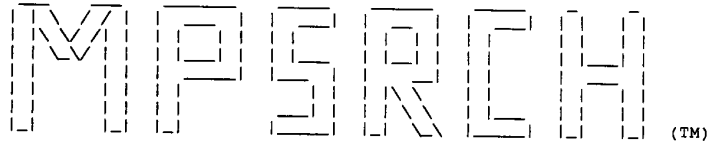
```
RESULT    9
ID   US-08-828-832-4      STANDARD;      PRT;    169 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 4, Application US/08828832
XX
CC    Sequence 4, Application US/08828832
CC    Patent No. 5827711
CC     GENERAL INFORMATION:
CC       APPLICANT:  Lal, Preeti
CC       APPLICANT:  Shah, Purvi
CC       TITLE OF INVENTION:  NOVEL SUCCINATE DEHYDROGENASE SUBUNIT
CC       NUMBER OF SEQUENCES:  4
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Incyte Pharmaceuticals, Inc.
CC         STREET:  3174 Porter Drive
CC         CITY:  Palo Alto
CC         STATE:  CA
CC         COUNTRY:  USA
CC         ZIP:  94304
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Diskette
CC         COMPUTER:  IBM Compatible
CC         OPERATING SYSTEM:  DOS
CC         SOFTWARE:  FastSEQ for Windows Version 2.0
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/828,832
CC         FILING DATE:  Herewith
CC         CLASSIFICATION:  435
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  Billings, Lucy J.
CC         REGISTRATION NUMBER:  36,749
CC         REFERENCE/DOCKET NUMBER:  PF-0250 US
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  415-855-0555
CC         TELEFAX:  415-845-4166
CC         TELEX:
CC     INFORMATION FOR SEQ ID NO:  4:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  169 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  linear
CC       IMMEDIATE SOURCE:
CC         LIBRARY:  GenBank
CC         CLONE:  1705529
SQ   SEQUENCE   169 AA; 18389 MW; 161800 CN;

  Query Match           14.3%;  Score 80;  DB 2;  Length 169;
  Best Local Similarity  30.8%;  Pred. No. 4.43e+01;
  Matches    16;  Conservative   18;  Mismatches  15;  Indels   3;  Gaps   3;

Db      60 GWSLPMAMSICHRGTGIALSAGVSL-FGLS-ALLVPGSFESH-LEFVKSLCL 108
           | :||: : :   |:: : ::|::| : |  ::|  :|:||  ||::  |||
Qy       2 GSGLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCL 53


RESULT   10
ID   US-07-817-920-6      STANDARD;      PRT;    390 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/07817920
XX
CC    Sequence 6, Application US/07817920
CC    Patent No. 5360735
CC     GENERAL INFORMATION:
CC       APPLICANT:  Weinshank, Richard L
CC       APPLICANT:  Branchek, Theresa
CC       APPLICANT:  Hartig, Paul R
CC       TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
CC       TITLE OF INVENTION:  USES THEREOF
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham
CC         STREET:  30 Rockefeller Plaza
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10112
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/07/817,920
CC         FILING DATE:  19920108
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White, John P
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  1795/39318
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-977-9550
CC         TELEFAX:  212-664-0525
CC         TELEX:  422523 COOP UI
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  390 amino acids
CC         TYPE:  AMINO ACID
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  N-terminal
CC       IMMEDIATE SOURCE:
CC         CLONE:  5-HT1DB
SQ   SEQUENCE   390 AA; 43656 MW; 849817 CN;

  Query Match           14.3%;  Score 80;  DB 1;  Length 390;
  Best Local Similarity  23.6%;  Pred. No. 4.43e+01;
  Matches    17;  Conservative   22;  Mismatches  32;  Indels   1;  Gaps   1;

Db      52 LVMLLALITLATTLSNAFVIATVYRTRKLHTPANYLIASLDVTDLLVSILVIPISTMYTV 111
           ||:||:|:  :   : :  :    :   | | ::|  : |:: : |  :| :| :|  |:
Qy       7 LVLLLTLLGSSHGTGPGMTLQLKLKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTL 65

Db     112 TDRWTLSQVVCD 123
             ::  :|||:
Qy      66 HHARSQHHVVCN 77


RESULT   11
ID   US-08-117-006-6      STANDARD;      PRT;    390 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08117006
XX
CC    Sequence 6, Application US/08117006
CC    Patent No. 5639652
CC     GENERAL INFORMATION:
CC       APPLICANT:  Weinshank, Richard L.
```

```
CC       APPLICANT:  Branchek, Theresa
CC       APPLICANT:  Hartig, Paul R.
CC       TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
CC       TITLE OF INVENTION:  USES THEREOF
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham
CC         STREET:  30 Rockefeller Plaza
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10112
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  WordPerfect, Version 5.1
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/117,006
CC         FILING DATE:
CC         CLASSIFICATION:  514
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White, John P
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  1795/39318
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-977-9550
CC         TELEFAX:  212-664-0525
CC         TELEX:  422523 COOP UI
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  390 amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  N-terminal
CC       IMMEDIATE SOURCE:
CC         CLONE:  5-HT1DB
SQ   SEQUENCE   390 AA; 43656 MW; 849817 CN;

  Query Match           14.3%;  Score 80;  DB 1;  Length 390;
  Best Local Similarity  23.6%;  Pred. No. 4.43e+01;
  Matches    17;  Conservative   22;  Mismatches  32;  Indels   1;  Gaps   1;

Db     52 LVMLLALITLATTLSNAFVIATVYRTRKLHTPANYLIASLDVTDLLVSILVIPISTMYTV 111
          ||:||:|:  :   : :  :    :   | | ::|  : |:: : |  :| :| :|  |:
Qy      7 LVLLLTLLGSSHGTGPGMTLQLKLKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTL 65

Db    112 TDRWTLSQVVCD 123
            ::  :|||:
Qy     66 HHARSQHHVVCN 77


RESULT   12
ID   PCT-US93-00149-6     STANDARD;     PRT;   390 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application PC/TUS9300149
XX
CC    Sequence 6, Application PC/TUS9300149
CC     GENERAL INFORMATION:
CC       APPLICANT:  Weinshank, Richard L.
CC       APPLICANT:  Branchek, Theresa
CC       APPLICANT:  Hartig, Paul R.
CC       TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
CC       TITLE OF INVENTION:  USES THEREOF
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham
CC         STREET:  30 Rockefeller Plaza
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10112
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  PCT/US93/00149
CC         FILING DATE:  19930108
CC         CLASSIFICATION:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White, John P
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  1795/39318
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-977-9550
CC         TELEFAX:  212-664-0525
CC         TELEX:  422523 COOP UI
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  390 amino acids
CC         TYPE:  AMINO ACID
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  N-terminal
CC       IMMEDIATE SOURCE:
CC         CLONE:  5-HT1DB
SQ   SEQUENCE   390 AA; 43656 MW; 849817 CN;

  Query Match           14.3%;  Score 80;  DB 3;  Length 390;
  Best Local Similarity  23.6%;  Pred. No. 4.43e+01;
  Matches    17;  Conservative   22;  Mismatches  32;  Indels   1;  Gaps   1;

Db     52 LVMLLALITLATTLSNAFVIATVYRTRKLHTPANYLIASLDVTDLLVSILVIPISTMYTV 111
          ||:||:|:  :   : :  :    :   | | ::|  : |:: : |  :| :| :|  |:
Qy      7 LVLLLTLLGSSHGTGPGMTLQLKLKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTL 65

Db    112 TDRWTLSQVVCD 123
            ::  :|||:
Qy     66 HHARSQHHVVCN 77


RESULT   13
ID   US-08-216-594-6     STANDARD;     PRT;   390 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08216594
XX
CC    Sequence 6, Application US/08216594
CC    Patent No. 5652113
CC     GENERAL INFORMATION:
CC       APPLICANT:  Weinshank, Richard L.
CC       APPLICANT:  Branchek, Theresa
CC       APPLICANT:  Hartig, Paul R.
CC       TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1F RECEPTOR
CC       TITLE OF INVENTION:  AND USES THEREOF
CC       NUMBER OF SEQUENCES:  9
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham
CC         STREET:  30 Rockefeller Plaza
CC         CITY:  New York
```

```
CC         STATE:  New York
CC         COUNTRY:  USA
CC         ZIP:  10112
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/216,594
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White, John P
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  1795/39318
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-977-9550
CC         TELEFAX:  212-664-0525
CC         TELEX:  422523 COOP UI
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  390 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  linear
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  N-terminal
CC       IMMEDIATE SOURCE:
CC         CLONE:  5-HT1DB
SQ   SEQUENCE   390 AA; 43656 MW; 849817 CN;

  Query Match           14.3%;  Score 80;  DB 1;  Length 390;
  Best Local Similarity  23.6%;  Pred. No. 4.43e+01;
  Matches   17;  Conservative   22;  Mismatches  32;  Indels   1;  Gaps   1;

Db      52 LVMLLALITLATTLSNAFVIATVYRTRKLHTPANYLIASLDVTDLLVSILVIPISTMYTV 111
           ||:||:|:  :   : :  :    :   | | ::|  : |:: : |  :| :| :|  |:
Qy       7 LVLLLTLLGSSHGTGPGMTLQLKLKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTL 65

Db     112 TDRWTLSQVVCD 123
             ::  :|||:
Qy      66 HHARSQHHVVCN 77


RESULT   14
ID   US-08-542-358-6      STANDARD;      PRT;   398 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08542358
XX
CC    Sequence 6, Application US/08542358
CC    Patent No. 5786155
CC     GENERAL INFORMATION:
CC       APPLICANT:  Weinshank, Richard L.
CC       APPLICANT:  Branchek, Theresa
CC       APPLICANT:  Hartig, Paul R.
CC       TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1E RECEPTOR AND USES THEREOF
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham LLP
CC         STREET:  1185 Avenue of the Americas
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10036
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/542,358
CC         FILING DATE:
CC         CLASSIFICATION:  435
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White, John P.
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  1795/39317-2Z/JPW/MAT
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-278-0400
CC         TELEFAX:  212-391-0525
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  398 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  N-terminal
SQ   SEQUENCE   398 AA; 44384 MW; 880684 CN;

  Query Match           14.3%;  Score 80;  DB 2;  Length 398;
  Best Local Similarity  23.6%;  Pred. No. 4.43e+01;
  Matches   17;  Conservative   22;  Mismatches  32;  Indels   1;  Gaps   1;

Db      52 LVMLLALITLATTLSNAFVIATVYRTRKLHTPANYLIASLDVTDLLVSILVIPISTMYTV 111
           ||:||:|:  :   : :  :    :   | | ::|  : |:: : |  :| :| :|  |:
Qy       7 LVLLLTLLGSSHGTGPGMTLQLKLKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTL 65

Db     112 TDRWTLSQVVCD 123
             ::  :|||:
Qy      66 HHARSQHHVVCN 77


RESULT   15
ID   US-08-370-542-6      STANDARD;      PRT;   398 AA.
XX
AC   xxxxxx
XX
DT
XX
DE   Sequence 6, Application US/08370542
XX
CC    Sequence 6, Application US/08370542
CC    Patent No. 5476782
CC     GENERAL INFORMATION:
CC       APPLICANT:  Weinshank, Richard L.
CC       APPLICANT:  Branchek, Theresa
CC       APPLICANT:  Hartig, Paul R.
CC       TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1E RECEPTOR AND
CC       TITLE OF INVENTION:  USES THEREOF
CC       NUMBER OF SEQUENCES:  12
CC       CORRESPONDENCE ADDRESS:
CC         ADDRESSEE:  Cooper & Dunham
CC         STREET:  30 Rockefeller Plaza
CC         CITY:  New York
CC         STATE:  New York
CC         COUNTRY:  U.S.A.
CC         ZIP:  10112
CC       COMPUTER READABLE FORM:
CC         MEDIUM TYPE:  Floppy disk
CC         COMPUTER:  IBM PC compatible
CC         OPERATING SYSTEM:  PC-DOS/MS-DOS
CC         SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC       CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/370,542
CC         FILING DATE:
CC         CLASSIFICATION:  435
```

```
CC       PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER:  US/08/194,113
CC         FILING DATE:
CC         APPLICATION NUMBER:  US/07/803,626
CC         FILING DATE:
CC       ATTORNEY/AGENT INFORMATION:
CC         NAME:  White, John P.
CC         REGISTRATION NUMBER:  28,678
CC         REFERENCE/DOCKET NUMBER:  1795/39317
CC       TELECOMMUNICATION INFORMATION:
CC         TELEPHONE:  212-977-9550
CC         TELEFAX:  212-664-0525
CC         TELEX:  422523 COOP UI
CC     INFORMATION FOR SEQ ID NO:  6:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  398 amino acids
CC         TYPE:  amino acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       HYPOTHETICAL:  NO
CC       ANTI-SENSE:  NO
CC       FRAGMENT TYPE:  N-terminal
SQ   SEQUENCE   398 AA; 44384 MW; 880684 CN;

  Query Match           14.3%;  Score 80;  DB 1;  Length 398;
  Best Local Similarity  23.6%;  Pred. No. 4.43e+01;
  Matches    17;  Conservative   22;  Mismatches  32;  Indels   1;  Gaps   1;

Db     52 LVMLLALITLATTLSNAFVIATVYRTRKLHTPANYLIASLDVTDLLVSILVIPISTMYTV 111
           ||:||:|:  :   : :  :    :   | | ::|  : |:: : |  :| :| :|  |:
Qy      7 LVLLLTLLGSSHGTGPGMTLQLKLKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTL 65

Db    112 TDRWTLSQVVCD 123
            ::  :|||:
Qy     66 HHARSQHHVVCN 77


Search completed: Fri Oct 22 18:45:24 1999
Job time : 9 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 18:42:40 1999; MasPar time 5.89 Seconds
530.469 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect Score: 558
Sequence: 1 MGSGLPLVLLLTLLGSSHGT.....SGTSVTLHHARSQHHVVCNT 78

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 39.113; Variance 85.189; scale 0.459

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	101	18.1	501	2	C70384	protein export membra	7.90e-02
2	93	16.7	382	2	S63638	ubiquinol--cytochrome	8.03e-01
3	92	16.5	712	2	A48156	translation regulator	1.07e+00
4	91	16.3	252	2	JC5237	osmotin-like protein	1.41e+00
5	89	15.9	879	2	S73757	hypothetical protein	2.46e+00
6	89	15.9	3005	2	S33642	homeotic protein zfh-	2.46e+00
7	88	15.8	1822	2	S44849	K12H4.8 protein - Cae	3.23e+00
8	87	15.6	143	2	S43071	hypothetical protein	4.25e+00
9	87	15.6	396	2	S58161	probable translation	4.25e+00
10	86	15.4	272	2	S39641	flagellar motor appar	5.57e+00
11	85	15.2	315	2	H64082	hypothetical protein	7.28e+00
12	84	15.1	441	2	S74800	dihydroorotase (EC 3.	9.51e+00
13	84	15.1	454	2	T03130	probable tyrosine kin	9.51e+00
14	84	15.1	509	2	S76731	hypothetical protein	9.51e+00
15	83	14.9	378	2	D70324	hypothetical protein	1.24e+01
16	83	14.9	401	1	YTECT0	tetracycline resistan	1.24e+01
17	83	14.9	469	2	G65058	hypothetical protein	1.24e+01
18	83	14.9	647	2	A37086	beta-galactosidase (E	1.24e+01
19	82	14.7	56	2	S53000	mitotic-specific cycl	1.61e+01
20	82	14.7	110	2	S23368	T-cell receptor alpha	1.61e+01
21	82	14.7	110	2	S22896	T-cell receptor alpha	1.61e+01
22	82	14.7	152	2	S58125	T-cell antigen recept	1.61e+01
23	82	14.7	155	2	S58178	T cell antigen recept	1.61e+01
24	82	14.7	372	2	A46138	invasion protein invE	1.61e+01
25	82	14.7	946	3	T00024	ent-Kaurene synthase	1.61e+01
26	81	14.5	253	2	S70170	motA protein - Rhodob	2.09e+01
27	81	14.5	261	2	I52518	sperm acrosome antige	2.09e+01
28	81	14.5	287	2	S71192	mitosis-specific cycl	2.09e+01
29	81	14.5	346	2	JC5715	G protein-coupled rec	2.09e+01
30	81	14.5	346	2	JC5716	G protein-coupled rec	2.09e+01
31	81	14.5	387	1	A48326	ubiquinol--cytochrome	2.09e+01
32	81	14.5	469	2	H70626	probable narK3 protei	2.09e+01
33	81	14.5	474	2	T00943	hypothetical protein	2.09e+01
34	81	14.5	482	2	B35843	lipopolysaccharide-bi	2.09e+01
35	81	14.5	1219	2	I61713	co-repressor protein	2.09e+01
36	81	14.5	1229	2	A56068	co-repressor protein	2.09e+01
37	80	14.3	132	2	A45159	succinate dehydrogena	2.70e+01
38	80	14.3	141	2	G71079	hypothetical protein	2.70e+01
39	80	14.3	191	2	H71370	hypothetical protein	2.70e+01
40	80	14.3	331	2	D70431	oligopeptide transpor	2.70e+01
41	80	14.3	387	1	CBASN	ubiquinol--cytochrome	2.70e+01
42	80	14.3	402	2	S23860	chloramphenicol resis	2.70e+01
43	80	14.3	741	2	I48694	probable transcriptio	2.70e+01
44	80	14.3	742	2	A49672	transcription factor	2.70e+01
45	80	14.3	772	2	A55004	transcription factor	2.70e+01

ALIGNMENTS

RESULT 1
ENTRY C70384 #type complete
TITLE protein export membrane protein SecD - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 12-Feb-1999
ACCESSIONS C70384
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MUID:98196666
#accession C70384
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-501 ##label AQF
##cross-references GB:AE000716; NID:g2983478; PID:g2983481; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene secD
CLASSIFICATION #superfamily protein export membrane protein secD
SUMMARY #length 501 #molecular-weight 55459 #checksum 2250

Query Match 18.1%; Score 101; DB 2; Length 501;
Best Local Similarity 38.3%; Pred. No. 7.90e-02;
Matches 18; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

```
Db    454 VILFQF-GSGPVKGFATTLALGTIASFISNVYYAKVFLDLLNSLKIL 499
          |:|: : ||:   | : || |    ||::| |   ||:||: | :|
Qy      8 VLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLL 54
```

RESULT 2
ENTRY S63638 #type complete
TITLE ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome b - Allomyces macrogynus mitochondrion
ALTERNATE_NAMES apocytochrome b
ORGANISM #formal_name mitochondrion Allomyces macrogynus
DATE 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 08-Sep-1997
ACCESSIONS S63638
REFERENCE S63635

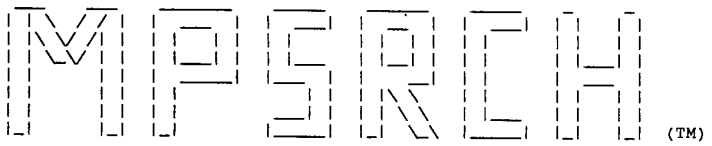
```
   #authors      Paquin, B.; Lang, B.F.
   #journal      J. Mol. Biol. (1996) 255:688-701
   #title        The mitochondrial DNA of Allomyces macrogynus: the complete
                   genomic sequence from an ancestral fungus.
   #cross-references MUID:96226032
   #accession    S63638
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-382 ##label PAQ
      ##cross-references EMBL:U41288; NID:g1236403; PID:g1236404
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, November 1995
GENETICS
   #gene         cob
   #genome       mitochondrion
   #introns      67/3; 137/3; 143/3; 164/1; 200/2; 252/3
CLASSIFICATION  #superfamily cytochrome b; cytochrome b homology; cytochrome
                  b6 homology; plastoquinol--plastocyanin reductase 17K
                  protein homology
KEYWORDS        mitochondrion; oxidoreductase
FEATURE
   10-340            #domain cytochrome b homology #label CBH\
   10-210            #domain cytochrome b6 homology #label CB6\
   222-340           #domain plastoquinol--plastocyanin reductase 17K protein
                       homology #label 17K
SUMMARY         #length 382  #molecular-weight 43467  #checksum 6973

  Query Match           16.7%;  Score 93;  DB 2;  Length 382;
  Best Local Similarity  33.3%;  Pred. No. 8.03e-01;
  Matches    17;  Conservative   15;  Mismatches  16;  Indels   3;  Gaps   3;

Db       8 PVLSLANSFLIDSPLPSNITYLWNFGSLLG-LCLVIQIVTGVTLAMHYAPS 57
           | ::|    | :| | :| :|  :|  ||  |||:::: :| ::::| | |
Qy      22 PGMTLQLK-LKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTLHHARS 70


RESULT    3
ENTRY           A48156     #type complete
TITLE           translation regulator GCD6 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES guanine nucleotide exchange factor chain GCD6; protein
                  YD8142B.03; protein YDR211w; translation initiation factor
                  eIF-2B homolog
ORGANISM        #formal_name Saccharomyces cerevisiae
DATE            28-May-1993 #sequence_revision 03-May-1994 #text_change
                  06-Feb-1998
ACCESSIONS      A48156; S61578; S30776
REFERENCE       A48156
   #authors     Bushman, J.L.; Asuru, A.I.; Matts, R.L.; Hinnebusch, A.G.
   #journal     Mol. Cell. Biol. (1993) 13:1920-1932
   #title       Evidence that GCD6 and GCD7, translational regulators of
                  GCN4, are subunits of the guanine nucleotide exchange
                  factor for eIF-2 in Saccharomyces cerevisiae.
   #cross-references MUID:93180841
   #accession   A48156
      ##molecule_type DNA
      ##residues      1-712 ##label BUS
      ##cross-references EMBL:L07115; NID:g171572; PID:g171574
      ##note          sequence extracted from NCBI backbone (NCBIN:126018,
                       NCBIP:126021)
REFERENCE       S61576
   #authors     Oliver, K.; Harris, D.
   #submission  submitted to the EMBL Data Library, December 1995
   #accession   S61578
      ##molecule_type DNA
      ##residues      1-712 ##label OLI
      ##cross-references EMBL:Z68195; NID:g1122341; PID:e213795; PID:g1122344;
                         MIPS:YDR211w
      ##experimental_source strain AB972
GENETICS
   #gene         SGD:GCD6
      ##cross-references SGD:S0002619; MIPS:YDR211w
   #map_position 4R
KEYWORDS        translation regulation
SUMMARY         #length 712  #molecular-weight 81160  #checksum 142

  Query Match           16.5%;  Score 92;  DB 2;  Length 712;
  Best Local Similarity  39.5%;  Pred. No. 1.07e+00;
  Matches    17;  Conservative    8;  Mismatches  14;  Indels   4;  Gaps   4;

Db      12 LGN-HGKNSDMDVEDRLQAVVLTDS-YETRFMPLTAVKPRCLL 52
           ||: ||  : | :: :|    ||:| ||: |: |   |  |||
Qy      14 LGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLE-KL-CLL 54


RESULT    4
ENTRY           JC5237     #type complete
TITLE           osmotin-like protein precursor - tomato
ORGANISM        #formal_name Lycopersicon esculentum #common_name tomato
DATE            13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change
                  13-Nov-1998
ACCESSIONS      JC5237
REFERENCE       JC5237
   #authors     Chen, R.; Wang, F.; Smith, A.G.
   #journal     Gene (1996) 179:301-302
   #title       A flower-specific gene encoding an osmotin-like protein from
                  Lycopersicon esculentum.
   #cross-references MUID:97128324
   #contents    flower
   #accession   JC5237
      ##status        preliminary; nucleic acid sequence not shown
      ##molecule_type mRNA
      ##residues      1-252 ##label CHE
      ##cross-references GB:L76632; NID:g1220536; PID:g1220537
COMMENT     This protein is rich in cysteine content and play a part of the
              flowers defense mechanisms against pathogens.
FEATURE
   1-23              #domain signal sequence #status predicted #label SIG\
   24-252            #product osmotin-like protein #status predicted #label
                       MAT
SUMMARY         #length 252  #molecular-weight 27265  #checksum 2939

  Query Match           16.3%;  Score 91;  DB 2;  Length 252;
  Best Local Similarity  54.2%;  Pred. No. 1.41e+00;
  Matches    13;  Conservative    4;  Mismatches   7;  Indels   0;  Gaps   0;

Db      10 LPLSLLFTLLSLSQSTNPNFILTL 33
           ||| ||:|||: |::| |   | |
Qy       5 LPLVLLLTLLGSSHGTGPGMTLQL 28


RESULT    5
ENTRY           S73757     #type complete
TITLE           hypothetical protein F11_orf879 - Mycoplasma pneumoniae (ATCC
                  29342) (SGC3)
ORGANISM        #formal_name Mycoplasma pneumoniae
   #variety     ATCC 29342
DATE            27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
                  17-Jul-1998
ACCESSIONS      S73757
REFERENCE       S73327
   #authors     Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li,
                  B.C.; Herrmann, R.
   #journal     Nucleic Acids Res. (1996) 24:4420-4449
   #title       Complete sequence analysis of the genome of the bacterium
                  Mycoplasma pneumoniae.
   #cross-references MUID:97105885
   #accession   S73757
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-879 ##label HIM
      ##cross-references EMBL:AE000042; GB:U00089; NID:g1674112; PID:g1674117
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, November 1996
GENETICS
   #genetic_code SGC3
```

```
SUMMARY          #length 879  #molecular-weight 101086  #checksum 8208

  Query Match           15.9%;  Score 89;  DB 2;  Length 879;
  Best Local Similarity  28.9%;  Pred. No. 2.46e+00;
  Matches    11;  Conservative   15;  Mismatches  11;  Indels   1;  Gaps   1;

Db    677 EGIPKDSNY-SSFVHLLDQKSLFLQLAKVSGIDINENK 713
          |::  :|:| |||: ||:   |:|:|:  ::: ::  :
Qy     32 ESFLTNSSYESSFLELLEKLCLLLHLPSGTSVTLHHAR 69


RESULT    6
ENTRY            S33642     #type complete
TITLE            homeotic protein zfh-2 - fruit fly (Drosophila melanogaster)
ORGANISM         #formal_name Drosophila melanogaster
DATE             20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
                   24-Sep-1998
ACCESSIONS       S33642; S27817
REFERENCE        S33641
   #authors      Fortini, M.E.; Lai, Z.; Rubin, G.M.
   #journal      Mech. Dev. (1991) 34:113-122
   #title        The Drosophila zfh-1 and zfh-2 genes encode novel proteins
                   containing both zinc-finger and homeodomain motifs.
   #accession    S33642
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-3005 ##label FOR
      ##cross-references EMBL:M63450; NID:g158822; PID:g158823
GENETICS
   #gene         zfh-2
      ##cross-references FlyBase:FBgn0004607
CLASSIFICATION   #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS         DNA binding; homeobox; nucleus; transcription regulation;
                   zinc finger
FEATURE
   1798-1854          #domain homeobox homology #label HOX1\
   2155-2211          #domain homeobox homology #label HOX2\
   2761-2817          #domain homeobox homology #label HOX3
SUMMARY          #length 3005  #molecular-weight 332056  #checksum 7516

  Query Match           15.9%;  Score 89;  DB 2;  Length 3005;
  Best Local Similarity  38.0%;  Pred. No. 2.46e+00;
  Matches    19;  Conservative   11;  Mismatches  17;  Indels   3;  Gaps   3;

Db   2165 QIKVLQEFFENNSYPKDSDLEYLSKL-LLLS-PRVIVVWFQNARQKQRKI 2212
          |:|: : |: |:||   | || | || |||  |    | :::||  :: :
Qy     27 QLKLKESFLTNSSY-ESSFLELLEKLCLLLHLPSGTSVTLHHARSQHHVV 75


RESULT    7
ENTRY            S44849     #type complete
TITLE            K12H4.8 protein - Caenorhabditis elegans
ORGANISM         #formal_name Caenorhabditis elegans
DATE             14-Sep-1994 #sequence_revision 12-May-1995 #text_change
                   05-Jun-1998
ACCESSIONS       S44849
REFERENCE        S44617
   #authors      Favello, A.D.
   #submission   submitted to the EMBL Data Library, May 1993
   #description  Sequence of the C. elegans cosmid K12H4.
   #accession    S44849
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-1822 ##label FAV
      ##cross-references EMBL:L14331; NID:g289702; PID:g289703
GENETICS
   #introns      17/3; 72/1; 116/3; 170/1; 216/3; 340/3; 401/3; 444/3; 512/2;
                   634/3; 694/3; 745/1; 855/2; 1053/1; 1113/3; 1215/3; 1321/1;
                   1396/2; 1458/3; 1507/3; 1563/2; 1645/3; 1677/3; 1804/2
CLASSIFICATION   #superfamily unassigned DEAD/H box helicases; DEAD/H box
                   helicase homology
KEYWORDS         ATP; P-loop
FEATURE
   33-496             #domain DEAD/H box helicase homology #label DEAD\
   33-40              #region nucleotide-binding motif A (P-loop)\
   141-146            #region nucleotide-binding motif B\
   145-148            #region DEXH motif
SUMMARY          #length 1822  #molecular-weight 208291  #checksum 7039

  Query Match           15.8%;  Score 88;  DB 2;  Length 1822;
  Best Local Similarity  31.4%;  Pred. No. 3.23e+00;
  Matches    16;  Conservative   17;  Mismatches  15;  Indels   3;  Gaps   3;

Db   1300 IGLGVSPCLLLTALTTSNAAD-GMSLERFETIGDSFLKFATTDYLYHTLLD 1349
          :| |::  |||| | :|:::  ||:|: :  : :|||  :: :  :  ||:
Qy      1 MGSGLPLVLLLTLLGSSHGTGPGMTLQ-L-KLKESFLTNSSYESSFLELLE 49


RESULT    8
ENTRY            S43071     #type complete
TITLE            hypothetical protein 5 - human herpesvirus 6
ORGANISM         #formal_name human herpesvirus 6
DATE             19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change
                   17-Mar-1999
ACCESSIONS       S43071
REFERENCE        S43067
   #authors      Thompson, J.; Choudhury, S.; Kashanchi, F.; Doniger, J.;
                   Berneman, Z.; Frenkel, N.; Rosenthal, L.J.
   #journal      Oncogene (1994) 9:1167-1175
   #title        A transforming fragment within the direct repeat region of
                   human herpesvirus type 6 that transactivates HIV-1.
   #cross-references MUID:94181269
   #accession    S43071
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-143 ##label THO
      ##cross-references EMBL:X73675; NID:g469952; PID:g469957
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, June 1993
SUMMARY          #length 143  #molecular-weight 13317  #checksum 6533

  Query Match           15.6%;  Score 87;  DB 2;  Length 143;
  Best Local Similarity  47.2%;  Pred. No. 4.25e+00;
  Matches    17;  Conservative    4;  Mismatches  14;  Indels   1;  Gaps   1;

Db     88 LGLGLGLGLGLGLGLGLGLGLGLGLGLAGGFL 123
          :| || | | | | || : | | |: | | |  :||
Qy      1 MGSGLPLVLLLTL-LGSSHGTGPGMTLQLKLKESFL 35


RESULT    9
ENTRY            S58161     #type complete
TITLE            probable translation releasing factor RF-1 - fission yeast
                   (Schizosaccharomyces pombe)
ALTERNATE_NAMES  SPAC2F7.17 protein
ORGANISM         #formal_name Schizosaccharomyces pombe
DATE             13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change
                   16-Oct-1998
ACCESSIONS       S58161
REFERENCE        S58145
   #authors      Gentles, S.; Churcher, C.M.
   #submission   submitted to the EMBL Data Library, July 1995
   #accession    S58161
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-396 ##label GEN
      ##cross-references EMBL:Z50142; NID:g1052783; PID:g1052800
GENETICS
   #introns      57/3; 90/3
CLASSIFICATION   #superfamily translation releasing factor
SUMMARY          #length 396  #molecular-weight 44954  #checksum 9362

  Query Match           15.6%;  Score 87;  DB 2;  Length 396;
  Best Local Similarity  45.0%;  Pred. No. 4.25e+00;
  Matches     9;  Conservative    8;  Mismatches   3;  Indels   0;  Gaps   0;
```

```
Db    281 LTHIPTGITVSMQDSRSQHQ 300
          | |:|:| :|::: :|||:
Qy     54 LLHLPSGTSVTLHHARSQHH 73


RESULT   10
ENTRY            S39641     #type complete
TITLE            flagellar motor apparatus homolog ytxD - Bacillus subtilis
ORGANISM         #formal_name Bacillus subtilis
DATE             08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change
                   24-Sep-1998
ACCESSIONS       S39641; B70003
REFERENCE        S39641
   #authors      Grundy, F.J.; Waters, D.A.; Takova, T.Y.; Henkin, T.M.
   #journal      Mol. Microbiol. (1993) 10:259-271
   #title        Identification of genes involved in utilization of acetate
                   and acetoin in Bacillus subtilis.
   #cross-references MUID:95020526
   #accession    S39641
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-272 ##label GRU
REFERENCE        A69580
   #authors      Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                   Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                   Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                   A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                   Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                   Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                   Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                   Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                   Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                   M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                   S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                   Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                   C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                   Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                   Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                   Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                   Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                   Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                   Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                   M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                   M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                   V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                   A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                   Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                   Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                   Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                   Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                   B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                   Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                   Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                   Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                   Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                   Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                   K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                   Yoshikawa, H.; Danchin, A.
   #journal      Nature (1997) 390:249-256
   #title        The complete genome sequence of the Gram-positive bacterium
                   Bacillus subtilis.
   #cross-references MUID:98044033
   #accession    B70003
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-272 ##label KUN
      ##cross-references GB:Z99119; GB:AL009126; NID:g2635411; PID:e1185846;
                          PID:g2635457
      ##experimental_source strain 168
GENETICS
   #gene         ytxD
CLASSIFICATION   #superfamily flagellar motor rotation protein
KEYWORDS         transmembrane protein
SUMMARY          #length 272  #molecular-weight 30143  #checksum 4312

  Query Match           15.4%;  Score 86;  DB 2;  Length 272;
  Best Local Similarity 34.6%;  Pred. No. 5.57e+00;
  Matches   18;  Conservative   12;  Mismatches  20;  Indels   2;  Gaps   1;

Db    164 IGTLVGLVLMLKNLNDPHMLGPNMAIALLTTLYGSLLANMVFNPIAAKLEEK 215
          :|: : |||:|  |  :|  || |:: |   |  |:|:|  :::    | ||
Qy      1 MGSGLPLVLLLTLLGSSHGTGPGMTLQL--KLKESFLTNSSYESSFLELLEK 50


RESULT   11
ENTRY            H64082     #type complete
TITLE            hypothetical protein HI0630 - Haemophilus influenzae (strain
                   Rd KW20)
ORGANISM         #formal_name Haemophilus influenzae
DATE             18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
                   04-Sep-1998
ACCESSIONS       H64082
REFERENCE        A64000
   #authors      Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
                   Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
                   Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
                   FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
                   Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
                   J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
                   M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
                   D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
                   J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
                   Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
   #journal      Science (1995) 269:496-512
   #title        Whole-genome random sequencing and assembly of Haemophilus
                   influenzae Rd.
   #cross-references MUID:95350630
   #accession    H64082
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-315 ##label TIGR
      ##cross-references GB:U32746; GB:L42023; NID:g1573626; PID:g1573627;
                          TIGR:HI0630
CLASSIFICATION   #superfamily conserved hypothetical protein HI0630
SUMMARY          #length 315  #molecular-weight 35906  #checksum 9329

  Query Match           15.2%;  Score 85;  DB 2;  Length 315;
  Best Local Similarity 29.5%;  Pred. No. 7.28e+00;
  Matches   13;  Conservative   18;  Mismatches  11;  Indels   2;  Gaps   2;

Db      9 TALSLSLLLSSIASAEELSAKQSLD-KMTQA-LDNLNYEIAFVQ 50
          ::|:| |||: ::|:   ::  :|: |: :: | | :|| :|::
Qy      3 SGLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLE 46


RESULT   12
ENTRY            S74800     #type complete
TITLE            dihydroorotase (EC 3.5.2.3) - Synechocystis sp. (strain PCC
                   6803)
ALTERNATE_NAMES  protein sll1018
ORGANISM         #formal_name Synechocystis sp.
   #variety      PCC 6803
DATE             25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                   21-Aug-1998
ACCESSIONS       S74800
REFERENCE        S74322
   #authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                   Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                   Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                   Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                   S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                   Yasuda, M.; Tabata, S.
   #journal      DNA Res. (1996) 3:109-136
   #title        Sequence analysis of the genome of the unicellular
                   cyanobacterium Synechocystis sp. PCC6803. II. Sequence
```

```
                determination of the entire genome and assignment of
                potential protein-coding regions.
   #cross-references MUID:97061201
   #accession    S74800
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-441 ##label KAN
      ##cross-references EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017684;
                        PID:g1652025
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, June 1996
GENETICS
   #gene          pyrC
KEYWORDS          hydrolase
SUMMARY           #length 441  #molecular-weight 48174  #checksum 3200

  Query Match           15.1%;  Score 84;  DB 2;  Length 441;
  Best Local Similarity  32.7%;  Pred. No. 9.51e+00;
  Matches    16;  Conservative   16;  Mismatches  15;  Indels   2;  Gaps   2;

Db    196 GKPLALVPLNLSLRGNGVMREGPLAIQLGLPPDPVMSEAAVIASLLELL 244
          |  |:|| | |:| |::      :::|| |  :: :::::  :|:||||
Qy      2 GSGLPLVLL-LTLLGSSHGTGPGMTLQLKLK-ESFLTNSSYESSFLELL 48


RESULT   13
ENTRY             T03130     #type complete
TITLE             probable tyrosine kinase - alcelaphine herpesvirus 1
ORGANISM          #formal_name alcelaphine herpesvirus 1
DATE              24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
                    24-Mar-1999
ACCESSIONS        T03130
REFERENCE         Z14840
   #authors       Ensser, A.; Pflanz, R.; Fleckenstein, B.
   #journal       J. Virol. (1997) 71:6517-6525
   #title         Primary structure of the alcelaphine herpesvirus 1 genome.
   #accession     T03130
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-454 ##label ENS
      ##cross-references EMBL:AF005370; NID:g2337967; PID:g2337998
SUMMARY           #length 454  #molecular-weight 50928  #checksum 4031

  Query Match           15.1%;  Score 84;  DB 2;  Length 454;
  Best Local Similarity  38.6%;  Pred. No. 9.51e+00;
  Matches    22;  Conservative   11;  Mismatches  20;  Indels   4;  Gaps   4;

Db    300 STLLQVLLL-VLARVMEQPNSLTYESWLRE-IEDDSS-DAYFLLLLGPKLVLLTHLS 353
          | |  |||| :|:       ::| :  |:| :  :|| :: || ||  || || ||:
Qy      3 SGLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLE-KLCLLLHLP 58


RESULT   14
ENTRY             S76731     #type complete
TITLE             hypothetical protein - Synechocystis sp. (strain PCC 6803)
ORGANISM          #formal_name Synechocystis sp.
   #variety       PCC 6803
DATE              25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
                    21-Aug-1998
ACCESSIONS        S76731
REFERENCE         S74322
   #authors       Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
                    Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.;
                    Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
                    Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
                    S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
                    Yasuda, M.; Tabata, S.
   #journal       DNA Res. (1996) 3:109-136
   #title         Sequence analysis of the genome of the unicellular
                    cyanobacterium Synechocystis sp. PCC6803. II. Sequence
                    determination of the entire genome and assignment of
                    potential protein-coding regions.
   #cross-references MUID:97061201
   #accession    S76731
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-509 ##label KAN
      ##cross-references EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019376;
                        PID:g1653732
      ##note          the nucleotide sequence was submitted to the EMBL Data
                       Library, June 1996
SUMMARY           #length 509  #molecular-weight 56720  #checksum 9647

  Query Match           15.1%;  Score 84;  DB 2;  Length 509;
  Best Local Similarity  40.0%;  Pred. No. 9.51e+00;
  Matches    14;  Conservative    7;  Mismatches  12;  Indels   2;  Gaps   2;

Db    402 GQGLPHVLLQDLLTEGEANQAEVNIQLTL-E-YLA 434
          | ||| |||  ||  : :  : : :|| | | :|:
Qy      2 GSGLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLT 36


RESULT   15
ENTRY             D70324     #type complete
TITLE             hypothetical protein aq_267 - Aquifex aeolicus
ORGANISM          #formal_name Aquifex aeolicus
DATE              08-May-1998 #sequence_revision 08-May-1998 #text_change
                    08-May-1998
ACCESSIONS        D70324
REFERENCE         A70300
   #authors       Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
                    Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
                    Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
                    J.M.; Olson, G.J.; Swanson, R.V.
   #journal       Nature (1998) 392:353-358
   #title         The complete genome of the hyperthermophilic bacterium
                    Aquifex aeolicus.
   #cross-references MUID:98196666
   #accession    D70324
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-378 ##label AQF
      ##cross-references GB:AE000681; NID:g2982963; PID:g2982971; GB:AE000657
      ##experimental_source strain VF5
GENETICS
   #gene          aq_267
SUMMARY           #length 378  #molecular-weight 43310  #checksum 6712

  Query Match           14.9%;  Score 83;  DB 2;  Length 378;
  Best Local Similarity  30.0%;  Pred. No. 1.24e+01;
  Matches    15;  Conservative   18;  Mismatches  13;  Indels   4;  Gaps   4;

Db    253 LPLGMLLGIFSASHAQALPFSLRLHLDLILYGFGAFTIF-GGMLHLLPRI 301
          ||| :|| ::::||: : | :: |:| |   :| : : : :: | || ::
Qy      5 LPLVLLLTLLGSSHGTG-P-GMTLQLKLKE-SFLTNSSYESSFLELLEKL 51


Search completed: Fri Oct 22 18:43:10 1999
Job time : 30 secs.
```

****



****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 18:43:27 1999; MasPar time 4.02 Seconds
548.762 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect Score: 558
Sequence: 1 MGSGLPLVLLLTLLGSSHGT.....SGTSVTLHHARSQHHVVCNT 78

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 40.060; Variance 77.694; scale 0.516

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	92	16.5	712	1	E2BE_YEAST	TRANSLATION INITIATION	3.33e-01
2	91	16.3	252	1	OLP1_LYCES	OSMOTIN-LIKE PROTEIN P	4.53e-01
3	89	15.9	3005	1	ZFH2_DROME	ZINC-FINGER PROTEIN 2	8.33e-01
4	88	15.8	1822	1	YM68_CAEEL	HYPOTHETICAL HELICASE	1.13e+00
5	87	15.6	396	1	RF1M_SCHPO	PUTATIVE MITOCHONDRIAL	1.52e+00
6	86	15.4	272	1	YTXD_BACSU	HYPOTHETICAL 30.1 KD P	2.04e+00
7	85	15.2	315	1	RSEB_HAEIN	SIGMA-E FACTOR REGULAT	2.74e+00
8	84	15.1	191	1	YDB3_SCHPO	HYPOTHETICAL 21.1 KD P	3.67e+00
9	83	14.9	401	1	TCR2_ECOLI	TETRACYCLINE RESISTANC	4.90e+00
10	83	14.9	445	1	YGCS_ECOLI	HYPOTHETICAL METABOLIT	4.90e+00
11	83	14.9	647	1	BGAL_MOUSE	BETA-GALACTOSIDASE PRE	4.90e+00
12	83	14.9	1233	1	NME3_HUMAN	GLUTAMATE [NMDA] RECEP	4.90e+00
13	82	14.7	372	1	INVE_SALTY	INVASION PROTEIN INVE.	6.52e+00
14	82	14.7	385	1	CYB_ASPNG	CYTOCHROME B (EC 1.10.	6.52e+00
15	81	14.5	253	1	MOTA_RHOSH	CHEMOTAXIS MOTA PROTEI	8.66e+00
16	81	14.5	261	1	ASPX_MOUSE	ACROSOMAL PROTEIN SP-1	8.66e+00
17	81	14.5	346	1	GP41_HUMAN	PUTATIVE G PROTEIN-COU	8.66e+00
18	81	14.5	346	1	GP42_HUMAN	PUTATIVE G PROTEIN-COU	8.66e+00
19	81	14.5	387	1	CYB_PODAN	CYTOCHROME B (EC 1.10.	8.66e+00
20	81	14.5	482	1	LBP_RABIT	LIPOPOLYSACCHARIDE-BIN	8.66e+00
21	80	14.3	169	1	C560_BOVIN	SUCCINATE DEHYDROGENAS	1.15e+01
22	80	14.3	191	1	Y064_TREPA	HYPOTHETICAL PROTEIN T	1.15e+01
23	80	14.3	387	1	CYB_EMENI	CYTOCHROME B (EC 1.10.	1.15e+01
24	80	14.3	400	1	TCR8_PASMU	TETRACYCLINE RESISTANC	1.15e+01
25	80	14.3	402	1	OPDE_PSEAE	TRANSCRIPTION REGULATO	1.15e+01
26	80	14.3	411	1	VGLM_HSVBC	GLYCOPROTEIN M.	1.15e+01
27	80	14.3	488	1	MBI1_EMENI	MRNA MATURASE BI1 (COB	1.15e+01
28	80	14.3	713	1	CDGT_BACSP	CYCLOMALTODEXTRIN GLUC	1.15e+01
29	80	14.3	772	1	TF11_HUMAN	TRANSCRIPTION FACTOR 1	1.15e+01
30	79	14.2	259	1	ATP6_YEAST	ATP SYNTHASE A CHAIN P	1.52e+01
31	79	14.2	408	1	BTN1_YEAST	BTN1 PROTEIN.	1.52e+01
32	79	14.2	413	1	RF1M_YEAST	MITOCHONDRIAL PEPTIDE	1.52e+01
33	79	14.2	493	1	ACHE_MOUSE	ACETYLCHOLINE RECEPTOR	1.52e+01
34	79	14.2	505	1	ACHB_BOVIN	ACETYLCHOLINE RECEPTOR	1.52e+01
35	79	14.2	511	1	MVIN_ECOLI	VIRULENCE FACTOR MVIN	1.52e+01
36	79	14.2	662	1	GARP_HUMAN	GARP PROTEIN PRECURSOR	1.52e+01
37	79	14.2	1015	1	PTPX_HUMAN	PROTEIN-TYROSINE PHOSP	1.52e+01
38	78	14.0	101	1	GRO_CRIGR	GROWTH REGULATED PROTE	2.00e+01
39	78	14.0	267	1	YTXD_BACME	HYPOTHETICAL 29.3 KD P	2.00e+01
40	78	14.0	490	1	FA10_RABIT	COAGULATION FACTOR X P	2.00e+01
41	78	14.0	991	1	BMP1_MOUSE	BONE MORPHOGENETIC PRO	2.00e+01
42	78	14.0	1239	1	NME3_MOUSE	GLUTAMATE [NMDA] RECEP	2.00e+01
43	77	13.8	253	1	POMA_VIBAL	CHEMOTAXIS POMA PROTEI	2.62e+01
44	77	13.8	530	1	UD18_RAT	UDP-GLUCURONOSYLTRANSF	2.62e+01
45	77	13.8	1156	1	HTS_DROME	HU-LI TAI SHAO PROTEIN	2.62e+01

ALIGNMENTS

RESULT 1
ID E2BE_YEAST STANDARD; PRT; 712 AA.
AC P32501;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRANSLATION INITIATION FACTOR EIF-2B EPSILON SUBUNIT (EIF-2B GDP-GTP
DE EXCHANGE FACTOR) (GUANINE NUCLEOTIDE EXCHANGE FACTOR SUBUNIT GCD6)
DE (GCD COMPLEX SUBUNIT GCD6).
GN GCD6 OR TIF225 OR YDR211W OR YD8142.12 OR YD8142B.03.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93180841.
RA BUSHMAN J.L., ASURU A.I., MATTS R.L., HINNEBUSCH A.G.;
RT "Evidence that GCD6 and GCD7, translational regulators of GCN4, are
RT subunits of the guanine nucleotide exchange factor for eIF-2 in
RT Saccharomyces cerevisiae.";
RL MOL. CELL. BIOL. 13:1920-1932(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA OLIVER K., SHORE L., HARRIS D., BARRELL B.G., RAJANDREAM M.A.,
RA WALSH S.V.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- FUNCTION: SUBUNIT OF THE GUANINE NUCLEOTIDE EXCHANGE FACTOR FOR
CC EIF-2. REQUIRED TO REPRESS GCN4 TRANSLATION UNDER NONSTARVATION
CC CONDITIONS. GCD6 AND GCD7 REPRESS GCN4 EXPRESSION AT THE
CC TRANSLATIONAL LEVEL BY ENSURING THAT RIBOSOMES WHICH HAVE
CC TRANSLATED UORF1 WILL REINITIATE AT UORF2, -3, OR -4 AND THUS FAIL
CC TO REACH THE GCN4 START SITE.
CC -!- SUBUNIT: COMPLEX OF FIVE DIFFERENT SUBUNITS; ALPHA (GCN3), BETA
CC (GCD7), GAMMA (GCD1), DELTA (GCD2) AND EPSILON (GCD6).
CC -!- SIMILARITY: BELONGS TO THE EIF-2B GAMMA/EPSILON SUBUNITS FAMILY.
CC ---

CC ---
DR EMBL; L07115; G171574; -.
DR EMBL; Z68194; E223719; -.

```
DR   EMBL; Z68195; E213795; -.
DR   PIR; S30776; S30776.
DR   PIR; A48156; A48156.
DR   SGD; L0000674; GCD6.
DR   PFAM; PF00132; hexapep; 3.
KW   AMINO-ACID BIOSYNTHESIS; TRANSLATION REGULATION.
SQ   SEQUENCE   712 AA;  81161 MW;  5DAD189F CRC32;

  Query Match           16.5%;  Score 92;  DB 1;  Length 712;
  Best Local Similarity  39.5%;  Pred. No. 3.33e-01;
  Matches    17;  Conservative    8;  Mismatches  14;  Indels   4;  Gaps   4;

Db     12 LGN-HGKNSDMDVEDRLQAVVLTDS-YETRFMPLTAVKPRCLL 52
          ||: ||  : | :: :|    ||:| ||: |: |   |  |||
Qy     14 LGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLE-KL-CLL 54


RESULT    2
ID   OLP1_LYCES     STANDARD;      PRT;   252 AA.
AC   Q41350;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   OSMOTIN-LIKE PROTEIN PRECURSOR.
OS   LYCOPERSICON ESCULENTUM (TOMATO).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC   ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; SOLANUM.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. VF36;
RX   MEDLINE; 97128324.
RA   CHEN R., WANG F., SMITH A.G.;
RT   "A flower-specific gene encoding an osmotin-like protein from
RT   Lycopersicon esculentum.";
RL   GENE 179:301-302(1996).
CC   -!- SUBCELLULAR LOCATION: CELL WALL PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L76632; G1220537; -.
DR   PROSITE; PS00316; THAUMATIN; 1.
DR   PFAM; PF00314; thaumatin; 1.
DR   HSSP; P25871; 1AUN.
KW   CELL WALL; SIGNAL.
FT   SIGNAL       1     24
FT   CHAIN       25    252       OSMOTIN-LIKE PROTEIN.
SQ   SEQUENCE   252 AA;  27265 MW;  229EB542 CRC32;

  Query Match           16.3%;  Score 91;  DB 1;  Length 252;
  Best Local Similarity  54.2%;  Pred. No. 4.53e-01;
  Matches    13;  Conservative    4;  Mismatches   7;  Indels   0;  Gaps   0;

Db     10 LPLSLLFTLLSLSQSTNPNFILTL 33
          ||| ||:|||: |::| |   | |
Qy      5 LPLVLLLTLLGSSHGTGPGMTLQL 28


RESULT    3
ID   ZFH2_DROME     STANDARD;      PRT;  3005 AA.
AC   P28167;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PROTEIN 2).
GN   ZFH-2.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92001539.
RA   FORTINI M.E., LAI Z., RUBIN G.M.;
RT   "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
RT   containing both zinc-finger and homeodomain motifs.";
RL   MECH. DEV. 34:113-122(1991).
CC   -!- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
CC       NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- TISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.
CC   -!- SIMILARITY: CONTAINS THREE HOMEOBOX DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63450; G158823; -.
DR   PIR; S27817; S27817.
DR   PIR; S33642; S33642.
DR   FLYBASE; FBgn0004607; zfh2.
DR   PROSITE; PS00027; HOMEOBOX_1; 2.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
DR   PROSITE; PS50071; HOMEOBOX_2; 3.
DR   PFAM; PF00046; homeobox; 3.
DR   PFAM; PF00096; zf-C2H2; 12.
DR   HSSP; P15822; 4ZNF.
DR   TRANSFAC; T00920; -.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING; HOMEOBOX; NUCLEAR PROTEIN;
KW   REPEAT.
FT   ZN_FING     133    156       C2H2-TYPE.
FT   ZN_FING     559    582       C2H2-TYPE.
FT   ZN_FING     614    638       C2H2-TYPE.
FT   ZN_FING     732    756       C2H2-TYPE.
FT   ZN_FING     897    916       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     940    964       C2H2-TYPE.
FT   ZN_FING     999   1023       C2H2-TYPE.
FT   ZN_FING    1074   1098       C2H2-TYPE.
FT   ZN_FING    1210   1233       C2H2-TYPE.
FT   ZN_FING    1341   1365       C2H2-TYPE.
FT   ZN_FING    1438   1462       C2H2-TYPE.
FT   ZN_FING    1477   1500       C2H2-TYPE (DEGENERATE).
FT   ZN_FING    1513   1535       C2H2-TYPE.
FT   ZN_FING    1541   1564       C2H2-TYPE.
FT   DNA_BIND   1797   1856       HOMEOBOX 1.
FT   DNA_BIND   2154   2213       HOMEOBOX 2.
FT   ZN_FING    2234   2256       C2H2-TYPE.
FT   ZN_FING    2371   2393       C2H2-TYPE.
FT   DNA_BIND   2760   2819       HOMEOBOX 3.
SQ   SEQUENCE   3005 AA;  332056 MW;  8B4CC45F CRC32;

  Query Match           15.9%;  Score 89;  DB 1;  Length 3005;
  Best Local Similarity  38.0%;  Pred. No. 8.33e-01;
  Matches    19;  Conservative   11;  Mismatches  17;  Indels   3;  Gaps   3;

Db   2165 QIKVLQEFFENNSYPKDSDLEYLSKL-LLLS-PRVIVVWFQNARQKQRKI 2212
          |:|: : |: |:||   | || | || |||  |    | :::||  :: :
Qy     27 QLKLKESFLTNSSY-ESSFLELLEKLCLLLHLPSGTSVTLHHARSQHHVV 75


RESULT    4
ID   YM68_CAEEL     STANDARD;      PRT;  1822 AA.
AC   P34529;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
```

```
DE   HYPOTHETICAL HELICASE K12H4.8 IN CHROMOSOME III.
GN   K12H4.8.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
CC   -!- SIMILARITY: WITH OTHER ATP DEPENDENT HELICASES.
CC   -!- SIMILARITY: CONTAINS A RNASE III DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L14331; G289703; -.
DR   PIR; S44849; S44849.
DR   WORMPEP; K12H4.8; CE00273.
DR   PROSITE; PS00517; RIBONUCLEASE_III; 1.
DR   PFAM; PF00035; dsrm; 1.
DR   PFAM; PF00271; helicase_C; 1.
DR   PFAM; PF00636; Ribonuclease_3; 2.
KW   HYPOTHETICAL PROTEIN; HELICASE; ATP-BINDING; HYDROLASE; NUCLEASE;
KW   ENDONUCLEASE.
FT   NP_BIND      33     40       ATP (POTENTIAL).
FT   SITE        145    148       DECH BOX.
FT   DOMAIN     1554   1822       RNASE III DOMAIN.
SQ   SEQUENCE   1822 AA;  208291 MW;  4F8E56BE CRC32;

  Query Match           15.8%;  Score 88;  DB 1;  Length 1822;
  Best Local Similarity  31.4%;  Pred. No. 1.13e+00;
  Matches    16;  Conservative   17;  Mismatches  15;  Indels   3;  Gaps   3;

Db   1300 IGLGVSPCLLLTALTTSNAAD-GMSLERFETIGDSFLKFATTDYLYHTLLD 1349
          :| |::  |||| | :|:::  ||:|: :  : :|||  :: :  :  ||:
Qy      1 MGSGLPLVLLLTLLGSSHGTGPGMTLQ-L-KLKESFLTNSSYESSFLELLE 49


RESULT    5
ID   RF1M_SCHPO     STANDARD;      PRT;   396 AA.
AC   Q09691;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   PUTATIVE MITOCHONDRIAL PEPTIDE CHAIN RELEASE FACTOR PRECURSOR.
GN   SPAC2F7.17.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   GENTLES S., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (JUL-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
CC       FACTORS FAMILY. STRONG, TO YEAST MRF-1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z50142; G1052800; -.
DR   PROSITE; PS00745; RF_PROK_I; 1.
DR   PFAM; PF00472; RF-1; 1.
KW   HYPOTHETICAL PROTEIN; PROTEIN BIOSYNTHESIS; MITOCHONDRION;
KW   TRANSIT PEPTIDE.
FT   TRANSIT      1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN        ?    396       PUTATIVE MITOCHONDRIAL PEPTIDE CHAIN
FT                               RELEASE FACTOR.
SQ   SEQUENCE   396 AA;  44954 MW;  988689CB CRC32;

  Query Match           15.6%;  Score 87;  DB 1;  Length 396;
  Best Local Similarity  45.0%;  Pred. No. 1.52e+00;
  Matches     9;  Conservative    8;  Mismatches   3;  Indels   0;  Gaps   0;

Db    281 LTHIPTGITVSMQDSRSQHQ 300
          | |:|:| :|::: :||||:
Qy     54 LLHLPSGTSVTLHHARSQHH 73


RESULT    6
ID   YTXD_BACSU     STANDARD;      PRT;   272 AA.
AC   P39063;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 30.1 KD PROTEIN IN ACUC 5'REGION (ORFA).
GN   YTXD.
OS   BACILLUS SUBTILIS.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC   BACILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 95020526.
RA   GRUNDY F.J., WATERS D.A., TAKOVA T.Y., HENKIN T.M.;
RT   "Identification of genes involved in utilization of acetate and
RT   acetoin in Bacillus subtilis.";
RL   MOL. MICROBIOL. 10:259-271(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98048467.
RA   LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH S.D.;
RT   "Sequencing and functional annotation of the Bacillus subtilis genes
RT   in the 200 kb rrnB-dnaB region.";
RL   MICROBIOLOGY 143:3431-3441(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN SOME TRANSPORT FUNCTION.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MOTA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L17309; G348048; -.
DR   EMBL; AF008220; G2293222; -.
DR   EMBL; Z99119; E1185846; -.
DR   PIR; S39641; S39641.
```

```
DR   SUBTILIST; BG10365; YTXD.
DR   PROSITE; PS01307; MOTA; 1.
KW   HYPOTHETICAL PROTEIN; TRANSPORT; TRANSMEMBRANE.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM     38     58       POTENTIAL.
FT   TRANSMEM    154    174       POTENTIAL.
FT   TRANSMEM    188    208       POTENTIAL.
FT   DOMAIN      209    272       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   272 AA;  30143 MW;  E89FDF9C CRC32;

  Query Match           15.4%;  Score 86;  DB 1;  Length 272;
  Best Local Similarity  34.6%;  Pred. No. 2.04e+00;
  Matches    18;  Conservative   12;  Mismatches  20;  Indels   2;  Gaps   1;

Db    164 IGTLVGLVLMLKNLNDPHMLGPNMAIALLTTLYGSLLANMVFNPIAAKLEEK 215
          :|: : |||:|  |  :|  || |:: |   |  |:|:|  :::    | ||
Qy      1 MGSGLPLVLLLTLLGSSHGTGPGMTLQL--KLKESFLTNSSYESSFLELLEK 50


RESULT    7
ID   RSEB_HAEIN     STANDARD;      PRT;   315 AA.
AC   P44792;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   SIGMA-E FACTOR REGULATORY PROTEIN RSEB HOMOLOG PRECURSOR.
GN   RSEB OR HI0630.
OS   HAEMOPHILUS INFLUENZAE.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC   HAEMOPHILUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   SCIENCE 269:496-512(1995).
CC   -!- FUNCTION: SEEMS TO MODULATE THE ACTIVITY OF RPOE (SIGMA-E)
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: TO E.COLI RSEB AND P.AERUGINOSA MUCB.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32746; G1573627; -.
DR   TIGR; HI0630; -.
KW   PERIPLASMIC; SIGNAL.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    315       SIGMA-E FACTOR REGULATORY PROTEIN RSEB
FT                                HOMOLOG.
SQ   SEQUENCE   315 AA;  35906 MW;  80B1C11D CRC32;

  Query Match           15.2%;  Score 85;  DB 1;  Length 315;
  Best Local Similarity  29.5%;  Pred. No. 2.74e+00;
  Matches    13;  Conservative   18;  Mismatches  11;  Indels   2;  Gaps   2;

Db      9 TALSLSLLLSSIASAEELSAKQSLD-KMTQA-LDNLNYEIAFVQ 50
          ::|:| |||: ::|:   ::  :|: |: :: | | :|| :|::
Qy      3 SGLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLE 46


RESULT    8
ID   YDB3_SCHPO     STANDARD;      PRT;   191 AA.
AC   Q10356;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 21.1 KD PROTEIN C22E12.03C IN CHROMOSOME I.
GN   SPAC22E12.03C.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SIMILARITY: BELONGS TO THE THIJ / PFPI FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z70043; E228197; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   191 AA;  21078 MW;  D47EB47B CRC32;

  Query Match           15.1%;  Score 84;  DB 1;  Length 191;
  Best Local Similarity  33.3%;  Pred. No. 3.67e+00;
  Matches    14;  Conservative   11;  Mismatches  16;  Indels   1;  Gaps   1;

Db    148 PVVLEENLI-TSQGPGTAMLFGLKLLEQVASKDKYNAVYKSL 188
          |:||   |: :|:| | :| : ||| |   ::  |:: :  |
Qy      6 PLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLEL 47


RESULT    9
ID   TCR2_ECOLI     STANDARD;      PRT;   401 AA.
AC   P02980;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   TETRACYCLINE RESISTANCE PROTEIN, CLASS B (TETA(B)) (METAL-
DE   TETRACYCLINE/H+ ANTIPORTER).
GN   TETA.
OS   ESCHERICHIA COLI.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ESCHERICHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84109550.
RA   NGUYEN T.T., POSTLE K., BERTRAND K.P.;
RT   "Sequence homology between the tetracycline-resistance determinants
RT   of Tn10 and pBR322.";
RL   GENE 25:83-92(1983).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83143319.
RA   HILLEN W., SCHOLLMEIER K.;
RT   "Nucleotide sequence of the Tn10 encoded tetracycline resistance
RT   gene.";
RL   NUCLEIC ACIDS RES. 11:525-539(1983).
RN   [3]
RP   MUTAGENESIS OF HIS-257.
RX   MEDLINE; 91177844.
RA   YAMAGUCHI A., ADACHI K., AKASAKA T., ONO N., SAWAI T.;
```

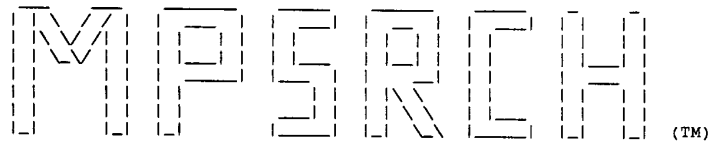
```
RT   "Metal-tetracycline/H+ antiporter of Escherichia coli encoded by a
RT   transposon Tn10. Histidine 257 plays an essential role in H+
RT   translocation.";
RL   J. BIOL. CHEM. 266:6045-6051(1991).
RN   [4]
RP   MUTAGENESIS OF 65-66.
RX   MEDLINE; 90368755.
RA   YAMAGUCHI A., ONO N., AKASAKA T., NOUMI T., SAWAI T.;
RT   "Metal-tetracycline/H+ antiporter of Escherichia coli encoded by a
RT   transposon, Tn10. The role of the conserved dipeptide, Ser65-Asp66,
RT   in tetracycline transport.";
RL   J. BIOL. CHEM. 265:15525-15530(1990).
CC   -!- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE
CC       EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE
CC       ACCUMULATION OF THE ANTIBIOTIC IN WHOLE CELLS. THIS PROTEIN
CC       FUNCTIONS AS A METAL-TETRACYCLINE/H+ ANTIPORTER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V00611; G43701; -.
DR   EMBL; J01830; G154847; -.
DR   PIR; A03507; YTECT0.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW   ANTIBIOTIC RESISTANCE; TRANSMEMBRANE; INNER MEMBRANE; TRANSPORT;
KW   SYMPORT; TRANSPOSABLE ELEMENT.
FT   TRANSMEM      6     26       POTENTIAL.
FT   TRANSMEM     42     62       POTENTIAL.
FT   TRANSMEM     74     94       POTENTIAL.
FT   TRANSMEM    102    122       POTENTIAL.
FT   TRANSMEM    131    151       POTENTIAL.
FT   TRANSMEM    159    179       POTENTIAL.
FT   TRANSMEM    214    234       POTENTIAL.
FT   TRANSMEM    244    265       POTENTIAL.
FT   TRANSMEM    277    297       POTENTIAL.
FT   TRANSMEM    298    318       POTENTIAL.
FT   TRANSMEM    336    356       POTENTIAL.
FT   TRANSMEM    363    383       POTENTIAL.
FT   MUTAGEN      65     65       S->C: ALMOST NO CHANGE IN ACTIVITY.
FT   MUTAGEN      65     65       S->A: NO CHANGE IN ACTIVITY.
FT   MUTAGEN      66     66       D->N: UNABLE TO EXTRUDE TETRACYCLINE.
FT   MUTAGEN      66     66       D->E: MODERATE RESISTANCE TO
FT                                TETRACYCLINE.
FT   MUTAGEN     257    257       H->E: NO H+ TRANSLOCATION.
FT   MUTAGEN     257    257       H->D: NO H+ TRANSLOCATION.
FT   CONFLICT    281    281       G -> E (IN REF. 2).
FT   CONFLICT    301    301       V -> D (IN REF. 2).
FT   CONFLICT    330    330       Q -> E (IN REF. 2).
FT   CONFLICT    354    354       A -> T (IN REF. 2).
SQ   SEQUENCE   401 AA;  43267 MW;  4823C395 CRC32;

  Query Match           14.9%;  Score 83;  DB 1;  Length 401;
  Best Local Similarity  31.0%;  Pred. No. 4.90e+00;
  Matches    18;  Conservative    13;  Mismatches  23;  Indels   4;  Gaps   4;

Db       1 MNSSTKIALVITLLDAM-GIGLIMPVLPTLLRE-FIASEDIANHFGVLLA-LYALMQV 55
           | |:  : |::||| :  | |  |  |   |:| |:::    : |  ||  |  |:::
Qy       1 MGSGLPLVLLLTLLGSSHGTGPGMT-LQLKLKESFLTNSSYESSFLELLEKLCLLLHL 57


RESULT   10
ID   YGCS_ECOLI     STANDARD;      PRT;   445 AA.
AC   Q46909;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN CYSJ-ENO INTERGENIC
DE   REGION.
GN   YGCS.
OS   ESCHERICHIA COLI.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   ESCHERICHIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   SCIENCE 277:1453-1474(1997).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29579; G882664; ALT_INIT.
DR   EMBL; AE000360; G1789130; ALT_INIT.
DR   ECOGENE; EG13126; YGCS.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR   PFAM; PF00083; sugar_tr; 1.
KW   HYPOTHETICAL PROTEIN; TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.
FT   TRANSMEM     23     43       POTENTIAL.
FT   TRANSMEM     57     77       POTENTIAL.
FT   TRANSMEM     86    106       POTENTIAL.
FT   TRANSMEM    115    135       POTENTIAL.
FT   TRANSMEM    143    163       POTENTIAL.
FT   TRANSMEM    176    196       POTENTIAL.
FT   TRANSMEM    254    274       POTENTIAL.
FT   TRANSMEM    287    307       POTENTIAL.
FT   TRANSMEM    312    332       POTENTIAL.
FT   TRANSMEM    338    358       POTENTIAL.
FT   TRANSMEM    370    390       POTENTIAL.
FT   TRANSMEM    401    421       POTENTIAL.
SQ   SEQUENCE   445 AA;  48234 MW;  BDD078EF CRC32;

  Query Match           14.9%;  Score 83;  DB 1;  Length 445;
  Best Local Similarity  38.6%;  Pred. No. 4.90e+00;
  Matches    17;  Conservative    9;  Mismatches  16;  Indels   2;  Gaps   2;

Db     296 GALLGLVLTH-LLAHRKFLLGSFLLLAATLVVMACLPSGSSLTL 338
             |  | | |   :|::  :  :||| |   | ::  ||||:|:||
Qy      23 GMTLQLKLKESFLTNSSYE-SSFLELLEKLCLLLHLPSGTSVTL 65


RESULT   11
ID   BGAL_MOUSE     STANDARD;      PRT;   647 AA.
AC   P23780;
DT   01-NOV-1991 (REL. 20, CREATED)
DT   01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE) (ACID BETA-
DE   GALACTOSIDASE).
GN   GLB1 OR GLB-1 OR BGL.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=BRAIN;
RX   MEDLINE; 91076843.
RA   NANBA E., SUZUKI K.;
RT   "Molecular cloning of mouse acid beta-galactosidase cDNA: sequence,
RT   expression of catalytic activity and comparison with the human
RT   enzyme.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 173:141-148(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2J;
RX   MEDLINE; 91298941.
RA   NANBA E., SUZUKI K.;
RT   "Organization of the mouse acid beta-galactosidase gene.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 178:158-164(1991).
CC   -!- FUNCTION: CLEAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM
CC       GANGLIOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCANS.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC       GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC   -!- SUBCELLULAR LOCATION: LYSOSOMAL.
CC   -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M57734; G192187; -.
DR   EMBL; M75122; G192185; -.
DR   EMBL; M75137; G192185; JOINED.
DR   EMBL; M75107; G192185; JOINED.
DR   EMBL; M75108; G192185; JOINED.
DR   EMBL; M75109; G192185; JOINED.
DR   EMBL; M75111; G192185; JOINED.
DR   EMBL; M75112; G192185; JOINED.
DR   EMBL; M75113; G192185; JOINED.
DR   EMBL; M75114; G192185; JOINED.
DR   EMBL; M75115; G192185; JOINED.
DR   EMBL; M75116; G192185; JOINED.
DR   EMBL; M75117; G192185; JOINED.
DR   EMBL; M75118; G192185; JOINED.
DR   EMBL; M75119; G192185; JOINED.
DR   EMBL; M75120; G192185; JOINED.
DR   EMBL; M75121; G192185; JOINED.
DR   PIR; A37086; A37086.
DR   MGD; MGI:88151; BGL.
DR   PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
DR   PFAM; PF01301; Glycosyl_hydr17; 1.
KW   HYDROLASE; GLYCOSIDASE; LYSOSOME; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     24
FT   PROPEP      25     29
FT   CHAIN       30    647       BETA-GALACTOSIDASE.
FT   ACT_SITE   189    189       PROTON DONOR (POTENTIAL).
FT   ACT_SITE   269    269       NUCLEOPHILE (POTENTIAL).
FT   CARBOHYD    27     27       POTENTIAL.
FT   CARBOHYD   248    248       POTENTIAL.
FT   CARBOHYD   500    500       POTENTIAL.
FT   CARBOHYD   504    504       POTENTIAL.
FT   CARBOHYD   510    510       POTENTIAL.
FT   CARBOHYD   544    544       POTENTIAL.
FT   CARBOHYD   557    557       POTENTIAL.
FT   CARBOHYD   617    617       POTENTIAL.
FT   CONFLICT   517    517       N -> D (IN REF. 2).
FT   CONFLICT   539    539       G -> R (IN REF. 2).
SQ   SEQUENCE   647 AA;  73121 MW;  15BCF158 CRC32;

  Query Match           14.9%;  Score 83;  DB 1;  Length 647;
  Best Local Similarity  37.8%;  Pred. No. 4.90e+00;
  Matches   14;  Conservative    8;  Mismatches  12;  Indels   3;  Gaps   3;

Db      10 LPLLALLQLLGAAHGIYNVTQRTFKLDYSRDRFLKDG 46
           |||: || |||::||       |: |   :: || ::
Qy       5 LPLVLLLTLLGSSHGT-G-PGMTLQLKL-KESFLTNS 38


RESULT   12
ID   NME3_HUMAN     STANDARD;      PRT;  1233 AA.
AC   Q14957;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 3 PRECURSOR (N-METHYL
DE   D-ASPARTATE RECEPTOR SUBTYPE 2C) (NR2C) (NMDAR2C).
GN   GRIN2C.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   LIN Y.J., BOVETTO S., CARVER J., GIORDANO T.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC       POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC       SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
CC   -!- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L76224; G1196449; -.
DR   MIM; 138254; -.
DR   PFAM; PF00060; lig_chan; 1.
KW   RECEPTOR; SIGNAL; TRANSMEMBRANE; POSTSYNAPTIC MEMBRANE; CALCIUM;
KW   IONIC CHANNEL; MAGNESIUM.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   1233       GLUTAMATE [NMDA] RECEPTOR SUBUNIT
FT                               EPSILON 3.
FT   TRANSMEM   554    574       1 (POTENTIAL).
FT   TRANSMEM   597    617       2 (POTENTIAL).
FT   TRANSMEM   627    647       3 (POTENTIAL).
FT   TRANSMEM   815    835       4 (POTENTIAL).
FT   SITE       612    612       FUNCTIONAL DETERMINANT OF NMDA
FT                               RECEPTORS (BY SIMILARITY).
FT   CARBOHYD    70     70       POTENTIAL.
FT   CARBOHYD   337    337       POTENTIAL.
FT   CARBOHYD   438    438       POTENTIAL.
FT   CARBOHYD'  539    539       POTENTIAL.
SQ   SEQUENCE   1233 AA;  134239 MW;  671F9981 CRC32;

  Query Match           14.9%;  Score 83;  DB 1;  Length 1233;
  Best Local Similarity  33.3%;  Pred. No. 4.90e+00;
  Matches   20;  Conservative   13;  Mismatches  24;  Indels   3;  Gaps   3;

Db      1 MGGALGPALLLTSLFGAWAGLGPGQGEQGMTVAVVFSSSGPPQAQFRARLTPQSFL-DLP 59
          ||:::|   |||| |:|:  | |||   | : :   | :::: :: |   |    :|  ||
Qy      1 MGSGLPLVLLLT-LLGSSHGTGPGMTLQ-LKLKESFLTNSSYESSFLELLEKLCLLLHLP 58


RESULT   13
ID   INVE_SALTY     STANDARD;      PRT;   372 AA.
AC   P35671;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INVASION PROTEIN INVE.
GN   INVE.
```

```
OS   SALMONELLA TYPHIMURIUM.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC   SALMONELLA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SR11 / SL1344;
RX   MEDLINE; 92335220.
RA   GINOCCHIO C., PACE J., GALAN J.E.;
RT   "Identification and molecular characterization of a Salmonella
RT   typhimurium gene involved in triggering the internalization of
RT   salmonellae into cultured epithelial cells.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 89:5976-5980(1992).
RN   [2]
RP   SEQUENCE OF 1-69 FROM N.A.
RC   STRAIN=SR11 / SL1344;
RX   MEDLINE; 95089692.
RA   KANIGA K., BOSSIO J.C., GALAN J.E.;
RT   "The Salmonella typhimurium invasion genes invF and invG encode
RT   homologues of the AraC and PulD family of proteins.";
RL   MOL. MICROBIOL. 13:555-568(1994).
RN   [3]
RP   SEQUENCE OF 1-5 FROM N.A.
RC   STRAIN=TML;
RA   LODGE J.M., AMIN I.I., DOUCE G.R., BROWN N.L., STEPHEN J.;
RL   SUBMITTED (SEP-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: INVOLVED IN THE TRIGGERING OF INTRACELLULAR EVENTS THAT
CC       LEAD TO MICROBIAL INTERNALIZATION OF THE INTESTINAL EPITHELIUM.
CC       THESE EVENTS INCLUDE INCREASE IN CALCIUM LEVEL, REDISTRIBUTION OF
CC       ACTIN MICROFILAMENTS, AND CHANGES IN THE NORMAL STRUCTURE OF THE
CC       MICROVILLI.
CC   -!- SIMILARITY: TO YERSINIA OUTER MEMBRANE PROTEIN YOPN (LCRE).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M90714; -; NOT_ANNOTATED_CDS.
DR   EMBL; U08280; G497226; -.
DR   EMBL; X75302; E86775; ALT_FRAME.
DR   STYGENE; SG10187; INVE.
KW   VIRULENCE.
SQ   SEQUENCE   372 AA;  42436 MW;  A345002E CRC32;

  Query Match           14.7%;  Score 82;  DB 1;  Length 372;
  Best Local Similarity  26.1%;  Pred. No. 6.52e+00;
  Matches    12;  Conservative   17;  Mismatches  16;  Indels   1;  Gaps   1;

Db    284 LLLMLSLLQQPHEVDSLLADIIGLNALLLSHKEH-ASFLQIFYQVC 328
          |:|:|:||  :|   : ::  : |:  :|::    :|||:::  :|
Qy      7 LVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLC 52


RESULT   14
ID   CYB_ASPNG      STANDARD;      PRT;   385 AA.
AC   Q33798;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CYTOCHROME B (EC 1.10.2.2).
GN   COB OR CYTB OR COBA.
OS   ASPERGILLUS NIGER.
OG   MITOCHONDRION.
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC   EUROTIALES; TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WU-2223L;
RA   NARUSAWA T., KANAYAMA S., KIRIMURA K., USAMI S.;
RT   "Nucleotide sequence of the apocytochrome b gene of Aspergillus
RT   niger WU-2223L.";
RL   SUBMITTED (JUL-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE OF 142-286 FROM N.A.
RC   STRAIN=IFM 5367, IFM 5368, IFM 40606, IFM 41398, IFM 41399, IFM 46897;
RX   MEDLINE; 99065785.
RA   WANG L., YOKOYAMA K., MIYAJI M., NISHIMURA K.;
RT   "The identification and phylogenetic relationship of pathogenic
RT   species of Aspergillus based on the mitochondrial cytochrome b gene.";
RL   MED. MYCOL. 36:153-164(1998).
CC   -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC       COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC       RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC       COUPLED TO ATP SYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q +
CC       2 FERROCYTOCHROME C.
CC   -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC       BOUND TO THE PROTEIN.
CC   -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC       CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D63375; G1000910; -.
DR   EMBL; AB000575; D1035119; -.
DR   EMBL; AB000576; D1035120; -.
DR   EMBL; AB000577; D1035121; -.
DR   EMBL; AB000578; D1035122; -.
DR   EMBL; AB000583; D1035127; -.
DR   EMBL; AB000597; D1035139; -.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PROSITE; PS00193; CYTOCHROME_B_QO; 1.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   ELECTRON TRANSPORT; MITOCHONDRION; RESPIRATORY CHAIN; TRANSMEMBRANE;
KW   HEME.
FT   METAL        82     82       IRON 1 (HEME B562 AXIAL LIGAND).
FT   METAL        96     96       IRON 2 (HEME B566 AXIAL LIGAND).
FT   METAL       183    183       IRON 2 (HEME B562 AXIAL LIGAND).
FT   METAL       197    197       IRON 1 (HEME B566 AXIAL LIGAND).
SQ   SEQUENCE   385 AA;  43078 MW;  17050CE1 CRC32;

  Query Match           14.7%;  Score 82;  DB 1;  Length 385;
  Best Local Similarity  34.3%;  Pred. No. 6.52e+00;
  Matches    12;  Conservative   11;  Mismatches  11;  Indels   1;  Gaps   1;

Db     24 ANISYLWNFGSLLA-LCLGIQIVTGVTLAMHYTPS 57
          :| ||  :|  ||  ||| ::: :| ::::| : |
Qy     36 TNSSYESSFLELLEKLCLLLHLPSGTSVTLHHARS 70


RESULT   15
ID   MOTA_RHOSH     STANDARD;      PRT;   253 AA.
AC   Q53174;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   CHEMOTAXIS MOTA PROTEIN (MOTILITY PROTEIN A).
GN   MOTA.
OS   RHODOBACTER SPHAEROIDES (RHODOPSEUDOMONAS SPHAEROIDES).
OC   BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;
OC   RHODOBACTER.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WS8;
RX   MEDLINE; 96123438.
RA   SHAH D.S.H., SOCKETT R.E.;
```

```
RT   "Analysis of the motA flagellar motor gene from Rhodobacter
RT   sphaeroides, a bacterium with a unidirectional, stop-start
RT   flagellum.";
RL   MOL. MICROBIOL. 17:961-969(1995).
CC   -!- FUNCTION: REQUIRED FOR ROTATION OF THE FLAGELLAR MOTOR. PROBABLE
CC       TRANSMEMBRANE PROTON CHANNEL (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE MOTA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85986; G758646; -.
DR   PROSITE; PS01307; MOTA; 1.
KW   CHEMOTAXIS; FLAGELLA; TRANSMEMBRANE; INNER MEMBRANE;
KW   FLAGELLAR ROTATION; HYDROGEN ION TRANSPORT.
FT   TRANSMEM      3     23       POTENTIAL.
FT   TRANSMEM     29     49       POTENTIAL.
FT   TRANSMEM    146    166       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
FT   DOMAIN      202    253       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   253 AA;  27194 MW;  F594C09C CRC32;

  Query Match           14.5%;  Score 81;  DB 1;  Length 253;
  Best Local Similarity  27.5%;  Pred. No. 8.66e+00;
  Matches    14;  Conservative   20;  Mismatches  14;  Indels   3;  Gaps   2;

Db    156 VGTLIGLVLMLGNMSDPKSIGPAMAVALLTTLYGALMANVIF-APILNKLE 205
          :|: : |||:|  :: : : ||:|:: |   |  ::::|  : :::|: ||
Qy      1 MGSGLPLVLLLTLLGSSHGTGPGMTLQL--KLKESFLTNSSYESSFLELLE 49


Search completed: Fri Oct 22 18:43:49 1999
Job time : 22 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Oct 22 18:44:06 1999; MasPar time 8.07 Seconds
527.481 Million cell updates/sec
Tabular output not generated.

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect Score: 558
Sequence: 1 MGSGLPLVLLLTLLGSSHGT.....SGTSVTLHHARSQHHVVCNT 78

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 38.815; Variance 83.141; scale 0.467

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	101	18.1	501	2	O67102	PROTEIN EXPORT MEMBRAN	8.20e-02
2	93	16.7	382	8	Q37395	APOCYTOCHROME B (EC 1.	8.75e-01
3	92	16.5	209	2	O68567	RESTRICTION ENDONUCLEA	1.17e+00
4	92	16.5	863	14	P89688	ENV POLYPROTEIN.	1.17e+00
5	89	15.9	879	2	P75377	F11_ORF879 PROTEIN.	2.74e+00
6	88	15.8	297	8	O03305	CYTOCHROME B (FRAGMENT	3.63e+00
7	87	15.6	143	14	Q69582	HERPESVIRUS TYPE 6 DNA	4.79e+00
8	86	15.4	322	4	Q14968	RHODOPSIN.	6.32e+00
9	86	15.4	562	5	Q23220	W08D2.3 PROTEIN.	6.32e+00
10	85	15.2	297	8	O03294	CYTOCHROME B (FRAGMENT	8.31e+00
11	85	15.2	297	8	O03295	CYTOCHROME B (FRAGMENT	8.31e+00
12	85	15.2	395	10	O23674	CHALCONE SYNTHASE HOMO	8.31e+00
13	85	15.2	448	13	O42453	SERPIN PRECURSOR.	8.31e+00
14	85	15.2	676	5	O45649	K03D7.3.	8.31e+00
15	84	15.1	371	4	O43190	PURINERGIC P2Y11 RECEP	1.09e+01
16	84	15.1	441	2	P72934	DIHYDROOROTASE.	1.09e+01
17	84	15.1	454	14	O36385	PUTATIVE TYROSINE KINA	1.09e+01
18	84	15.1	509	2	P74537	HYPOTHETICAL 56.7 KD P	1.09e+01
19	84	15.1	1016	5	O17485	AMINOPEPTIDASE.	1.09e+01
20	84	15.1	1016	5	O17484	AMINOPEPTIDASE.	1.09e+01
21	83	14.9	244	8	Q35487	CYTOCHROME B (FRAGMENT	1.43e+01
22	83	14.9	297	8	O03290	CYTOCHROME B (FRAGMENT	1.43e+01
23	83	14.9	378	2	O66624	HYPOTHETICAL 43.3 KD P	1.43e+01
24	83	14.9	385	8	O63916	CYTOCHROME B.	1.43e+01
25	83	14.9	711	5	O45825	T25E12.11 PROTEIN.	1.43e+01
26	83	14.9	1236	4	O15398	N-METHYL-D-ASPARTATE R	1.43e+01
27	82	14.7	135	4	O75609	INTEGRAL MEMBRANE PROT	1.87e+01
28	82	14.7	169	4	Q99643	INTEGRAL MEMBRANE PROT	1.87e+01
29	82	14.7	355	4	O75774	COL1A1 AND PDGFB FUSIO	1.87e+01
30	82	14.7	372	2	Q54027	INVASION PROTEIN.	1.87e+01
31	82	14.7	372	2	Q54032	INVASION PROTEIN.	1.87e+01
32	82	14.7	372	2	Q57178	INVASION PROTEIN.	1.87e+01
33	82	14.7	372	2	Q54029	INVASION PROTEIN.	1.87e+01
34	82	14.7	372	2	Q54031	INVASION PROTEIN.	1.87e+01
35	82	14.7	372	2	Q54037	INVASION PROTEIN.	1.87e+01
36	82	14.7	372	2	Q54036	INVASION PROTEIN.	1.87e+01
37	82	14.7	372	2	Q56052	INVASION PROTEIN.	1.87e+01
38	82	14.7	372	2	Q54034	INVASION PROTEIN.	1.87e+01
39	82	14.7	372	2	Q54028	INVASION PROTEIN.	1.87e+01
40	82	14.7	372	2	Q54025	INVASION PROTEIN.	1.87e+01
41	82	14.7	372	2	Q54030	INVASION PROTEIN.	1.87e+01
42	82	14.7	372	2	Q54026	INVASION PROTEIN.	1.87e+01
43	82	14.7	946	3	O13284	ENT-KAURENE SYNTHASE.	1.87e+01
44	82	14.7	1282	11	Q60520	PAIRED AMPHIPATHIC HEL	1.87e+01
45	82	14.7	5825	10	O82731	ORFA+B.	1.87e+01

ALIGNMENTS

RESULT 1
ID O67102 PRELIMINARY; PRT; 501 AA.
AC O67102;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PROTEIN EXPORT MEMBRANE PROTEIN SECD.
GN SECD.
OS AQUIFEX AEOLICUS.
OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; AE000716; G2983481; -.
SQ SEQUENCE 501 AA; 55459 MW; E67F690C CRC32;

Query Match 18.1%; Score 101; DB 2; Length 501;
Best Local Similarity 38.3%; Pred. No. 8.20e-02;
Matches 18; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

```
Db    454 VILFQF-GSGPVKGFATTLALGTIASFISNVYYAKVFLDLLNSLKIL 499
          |:|: : ||:   | : || |    ||::|  |   ||:||: | :|
Qy      8 VLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLL 54
```

RESULT 2
ID Q37395 PRELIMINARY; PRT; 382 AA.
AC Q37395;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

```
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   APOCYTOCHROME B (EC 1.10.2.2) (UBIQUINOL--CYTOCHROME C REDUCTASE)
DE   (CYTOCHROME BC1 COMPLEX).
GN   COB.
OS   ALLOMYCES MACROGYNUS.
OG   MITOCHONDRION.
OC   EUKARYOTA; FUNGI; CHYTRIDIOMYCOTA; CHYTRIDIOMYCETES; BLASTOCLADIALES;
OC   BLASTOCLADIACEAE; ALLOMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96226032.
RA   PAQUIN B., LANG B.F.;
RT   "The mitochondrial DNA of Allomyces macrogynus: the complete genomic
RT   sequence from an ancestral fungus.";
RL   J. MOL. BIOL. 255:688-701(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   PAQUIN B., LAFOREST M.J., LANG B.F.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q + 2
CC       FERROCYTOCHROME C.
DR   EMBL; U41288; G1236404; -.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   OXIDOREDUCTASE; MITOCHONDRION.
SQ   SEQUENCE   382 AA;  43467 MW;  A21E1AB1 CRC32;

  Query Match           16.7%;  Score 93;  DB 8;  Length 382;
  Best Local Similarity  33.3%;  Pred. No. 8.75e-01;
  Matches    17;  Conservative   15;  Mismatches  16;  Indels   3;  Gaps   3;

Db       8 PVLSLANSFLIDSPLPSNITYLWNFGSLLG-LCLVIQIVTGVTLAMHYAPS 57
           | ::|    | :| | :| :|  :|  ||  |||:::: :| ::::| | |
Qy      22 PGMTLQLK-LKESFL-TNSSYESSFLELLEKLCLLLHLPSGTSVTLHHARS 70


RESULT    3
ID   O68567      PRELIMINARY;      PRT;   209 AA.
AC   O68567;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE   RESTRICTION ENDONUCLEASE R.XBAI.
GN   XBAIR.
OS   XANTHOMONAS CAMPESTRIS.
OC   BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; LYSOBACTERALES;
OC   XANTHOMONAS GROUP; XANTHOMONAS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BADRII;
RA   ZHANG B.-H., WILSON G.G.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF051092; G2995645; -.
KW   ENDONUCLEASE.
SQ   SEQUENCE   209 AA;  23834 MW;  1EC63D0B CRC32;

  Query Match           16.5%;  Score 92;  DB 2;  Length 209;
  Best Local Similarity  28.2%;  Pred. No. 1.17e+00;
  Matches    20;  Conservative   21;  Mismatches  27;  Indels   3;  Gaps   2;

Db      31 GDDVSHYLIYRVLGVAQEEGRLIDVYQN-KGRFLYK--YAGSFLEAATKLCFKEAFPDSA 87
           |  ::  |:  :|| ::  |  : :  : |  || :  | :|||   |||:   :| ::
Qy       2 GSGLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLLLHLPSGT 61

Db      88 SLRLPNTQGQR 98
           |: | ::::|:
Qy      62 SVTLHHARSQH 72


RESULT    4
ID   P89688      PRELIMINARY;      PRT;   863 AA.
AC   P89688;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ENV POLYPROTEIN.
GN   ENV.
OS   FELINE IMMUNODEFICIENCY VIRUS (FIV).
OC   VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FIV-OMA;
RX   MEDLINE; 96013855.
RA   BARR M.C., ZOU L., HOLZSCHU D.L., PHILLIPS L., SCOTT F.W., CASEY J.W.,
RA   AVERY R.J.;
RT   "Isolation of a highly cytopathic lentivirus from a nondomestic
RT   cat.";
RL   J. VIROL. 69:7371-7374(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FIV-OMA;
RX   MEDLINE; 97177285.
RA   BARR M.C., ZOU L., LONG F., HOOSE W.A., AVERY R.J.;
RT   "Proviral organization and sequence analysis of feline
RT   immunodeficiency virus isolated from a Pallas' cat.";
RL   VIROLOGY 228:84-91(1997).
DR   EMBL; U56928; G1899040; -.
KW   POLYPROTEIN.
SQ   SEQUENCE   863 AA;  99281 MW;  6928E60F CRC32;

  Query Match           16.5%;  Score 92;  DB 14;  Length 863;
  Best Local Similarity  33.3%;  Pred. No. 1.17e+00;
  Matches    23;  Conservative   16;  Mismatches  27;  Indels   3;  Gaps   2;

Db     626 LGLTLALVTATTAGLIGTTTGTSA-LAVSLKLKEVMLQQSQINEATLGMLKILQRRLKQA 684
           :|  |:||   |  |:|:: ||:: ::: |||||  |  |  : : | :|  |   |  :
Qy       1 MGSGLPLVLLLT--LLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLLLHLP 58

Db     685 ERVILTLHQ 693
               :|||:
Qy      59 SGTSVTLHH 67


RESULT    5
ID   P75377      PRELIMINARY;      PRT;   879 AA.
AC   P75377;
DT   01-FEB-1997 (TREMBLREL. 02, CREATED)
DT   01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F11_ORF879 PROTEIN.
OS   MYCOPLASMA PNEUMONIAE.
OC   BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC   MYCOPLASMATACEAE; MYCOPLASMA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M129;
RX   MEDLINE; 97105885.
RA   HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.C.,
RA   HERRMANN R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   NUCLEIC ACIDS RES. 24:4420-4449(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M129;
RA   HIMMELREICH R., HILBERT H., LI B.-C.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AE000042; G1674117; -.
SQ   SEQUENCE   879 AA;  101086 MW;  94D217E0 CRC32;

  Query Match           15.9%;  Score 89;  DB 2;  Length 879;
  Best Local Similarity  28.9%;  Pred. No. 2.74e+00;
  Matches    11;  Conservative   15;  Mismatches  11;  Indels   1;  Gaps   1;

Db     677 EGIPKDSNY-SSFVHLLDQKSLFLQLAKVSGIDINENK 713
           |::  :|:| |||: ||:   |:|:|:  ::: ::  :
```

```
Qy      32 ESFLTNSSYESSFLELLEKLCLLLHLPSGTSVTLHHAR 69


RESULT    6
ID   O03305      PRELIMINARY;      PRT;   297 AA.
AC   O03305;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CYTOCHROME B (FRAGMENT).
OS   CARETTOCHELYS INSCULPTA (PITTED-SHELLED TURTLE).
OG   MITOCHONDRION.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TESTUDINES; CRYPTODIRA;
OC   TRIONYCHOIDEA; CARETTOCHELYIDAE; CARETTOCHELYS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SHAFFER H.B., MEYLAN P., MCKNIGHT M.L.;
RL   SYST. BIOL. 0:0-0(0).
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q + 2
CC       FERROCYTOCHROME C.
CC   -!- COFACTOR: TWO HEME GROUPS
CC       (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
CC       (BY SIMILARITY).
DR   EMBL; U81355; G2098653; -.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   MITOCHONDRION; ELECTRON TRANSPORT; RESPIRATORY CHAIN; TRANSMEMBRANE;
KW   HEME.
FT   NON_TER       1      1
FT   NON_TER     297    297
SQ   SEQUENCE   297 AA;  33587 MW;  7D5808B5 CRC32;

  Query Match           15.8%;  Score 88;  DB 8;  Length 297;
  Best Local Similarity  31.5%;  Pred. No. 3.63e+00;
  Matches    17;  Conservative   10;  Mismatches  25;  Indels   2;  Gaps   2;

Db    165 GLTIVHLLFLYETGSNNPTGLNSNMD-KIPFHPYFSYKDVF-GLILMLAILLNL 216
          || :| || |  ::     |:   :  | :|    ||   |  |: :| :||:|
Qy      4 GLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLLLHL 57


RESULT    7
ID   Q69582      PRELIMINARY;      PRT;   143 AA.
AC   Q69582;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   HERPESVIRUS TYPE 6 DNA.
OS   HUMAN HERPESVIRUS-6.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE; BETAHERPESVIRINAE;
OC   ROSEOLOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94181269.
RA   THOMPSON J., CHOUDHURY S., KASHANCHI F., DONIGER J., BERNEMAN Z.,
RA   FRENKEL N., ROSENTHAL L.J.;
RT   "A transforming fragment within the direct repeat region of human
RT   herpesvirus type 6 that transactivates HIV-1.";
RL   ONCOGENE 9:1167-1175(1994).
DR   EMBL; X73675; G469957; -.
SQ   SEQUENCE   143 AA;  13317 MW;  597857A6 CRC32;

  Query Match           15.6%;  Score 87;  DB 14;  Length 143;
  Best Local Similarity  47.2%;  Pred. No. 4.79e+00;
  Matches    17;  Conservative    4;  Mismatches  14;  Indels   1;  Gaps   1;

Db     88 LGLGLGLGLGLGLGLGLGLGLGLGLGLGLAGGFL 123
          :| || | | | | || : | | |: | | |  :||
Qy      1 MGSGLPLVLLLTL-LGSSHGTGPGMTLQLKLKESFL 35


RESULT    8
ID   Q14968      PRELIMINARY;      PRT;   322 AA.
AC   Q14968;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   RHODOPSIN.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92250505.
RA   KUNZ D., GERARD N.P., GERARD C.;
RT   "The human leukocyte platelet-activating factor receptor. cDNA
RT   cloning, cell surface expression, and construction of a novel
RT   epitope-bearing analog.";
RL   J. BIOL. CHEM. 267:9101-9106(1992).
DR   EMBL; M76676; G189270; -.
DR   PFAM; PF00001; 7tm_1; 2.
KW   GTP-BINDING.
SQ   SEQUENCE   322 AA;  33096 MW;  70F54EC8 CRC32;

  Query Match           15.4%;  Score 86;  DB 4;  Length 322;
  Best Local Similarity  34.5%;  Pred. No. 6.32e+00;
  Matches    19;  Conservative   13;  Mismatches  22;  Indels   1;  Gaps   1;

Db     95 ALVLLLIFLLSSLGNCAVMGVIVKHRQLRTVTNAFILS-LSLSDLLTALLCLPAA 148
          :||||| :| || |  : | : :| ::     :::  | | | : |  || ||::
Qy      6 PLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLLLHLPSG 60


RESULT    9
ID   Q23220      PRELIMINARY;      PRT;   562 AA.
AC   Q23220;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   W08D2.3 PROTEIN.
GN   W08D2.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SWINBURNE J., AINSCOUGH R.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z70271; E1350111; -.
SQ   SEQUENCE   562 AA;  63331 MW;  5272E400 CRC32;

  Query Match           15.4%;  Score 86;  DB 5;  Length 562;
  Best Local Similarity  31.5%;  Pred. No. 6.32e+00;
  Matches    23;  Conservative   19;  Mismatches  27;  Indels   4;  Gaps   4;

Db    122 IILLLCIVFIVFGTVLGISLSAVISN-FVVIFK-QQSFL-LLQFLHLIGAFGSVTAVMFY 178
          ::||| ::    ||  |::|   : : |:     : ||| ||: | |:  : | |:| :
Qy      7 LVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLLLHLPSGTSVTLH 66
```

```
Db    179 QSAPSEHQEDCNS 191
          : | |:|:  ||:
Qy     67 H-ARSQHHVVCNT 78


RESULT   10
ID   O03294      PRELIMINARY;      PRT;   297 AA.
AC   O03294;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CYTOCHROME B (FRAGMENT).
OS   HEOSEMYS SPINOSA.
OG   MITOCHONDRION.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TESTUDINES; CRYPTODIRA;
OC   TESTUDINOIDEA; BATAGURIDAE; HEOSEMYS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SHAFFER H.B., MEYLAN P., MCKNIGHT M.L.;
RL   SYST. BIOL. 0:0-0(0).
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q + 2
CC       FERROCYTOCHROME C.
CC   -!- COFACTOR: TWO HEME GROUPS
CC       (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
CC       (BY SIMILARITY).
DR   EMBL; U81362; G2098675; -.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   MITOCHONDRION; ELECTRON TRANSPORT; RESPIRATORY CHAIN; TRANSMEMBRANE;
KW   HEME.
FT   NON_TER       1      1
FT   NON_TER     297    297
SQ   SEQUENCE   297 AA;  33321 MW;  AD3A6749 CRC32;

  Query Match           15.2%;  Score 85;  DB 8;  Length 297;
  Best Local Similarity  32.1%;  Pred. No. 8.31e+00;
  Matches    18;  Conservative   11;  Mismatches  25;  Indels   2;  Gaps   2;

Db    164 AGLAIVHLLFLHETGSNNPTGLNSNAD-KIPFHPYFSYKD-LLGLILMLTLLLTLA 217
          :||::| || |  ::     |:      | :|    ||   :| |: :| ||| |:
Qy      3 SGLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLLLHLP 58


RESULT   11
ID   O03295      PRELIMINARY;      PRT;   297 AA.
AC   O03295;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CYTOCHROME B (FRAGMENT).
OS   TRACHEMYS SCRIPTA (RED-EARED SLIDER TURTLE).
OG   MITOCHONDRION.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TESTUDINES; CRYPTODIRA;
OC   TESTUDINOIDEA; EMYDIDAE; TRACHEMYS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SHAFFER H.B., MEYLAN P., MCKNIGHT M.L.;
RL   SYST. BIOL. 0:0-0(0).
CC   -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q + 2
CC       FERROCYTOCHROME C.
CC   -!- COFACTOR: TWO HEME GROUPS
CC       (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
CC       (BY SIMILARITY).
DR   EMBL; U81351; G2098693; -.
DR   PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   MITOCHONDRION; ELECTRON TRANSPORT; RESPIRATORY CHAIN; TRANSMEMBRANE;
KW   HEME.
FT   NON_TER       1      1
FT   NON_TER     297    297
SQ   SEQUENCE   297 AA;  33479 MW;  1BCBCFAC CRC32;

  Query Match           15.2%;  Score 85;  DB 8;  Length 297;
  Best Local Similarity  33.3%;  Pred. No. 8.31e+00;
  Matches    18;  Conservative    9;  Mismatches  25;  Indels   2;  Gaps   2;

Db    165 GLTLVHLLFLHETGSNNPTGLNSNVD-KIPFHPYFSYKD-LLGIILMLTLLLTL 216
          || || || |  ::     |:   :  | :|    ||   :| :: :| ||| |
Qy      4 GLPLVLLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKLCLLLHL 57


RESULT   12
ID   O23674      PRELIMINARY;      PRT;   395 AA.
AC   O23674;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CHALCONE SYNTHASE HOMOLOG.
GN   T7I23.4.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FEDERSPIEL N.A., CONWAY A.B., CONWAY A.R., DAVIS K., BRENDEL V.,
RA   PALM C.J., AU M., ARAUJO R., CHUNG E., KURTZ D.B., BUEHLER E.,
RA   DEWAR K., FENG J., KIM C., LI Y., SHINN P., SUN H., OJI O.,
RA   OSBORNE B., SHEN Y.K., TORIUMI M., VYOTSKAIA V., YU G., THEOLOGIS A.,
RA   ECKER J., DAVIS R.W.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U89959; G2317904; -.
DR   PFAM; PF00195; Chal_stil_synt; 1.
SQ   SEQUENCE   395 AA;  43617 MW;  B5894470 CRC32;

  Query Match           15.2%;  Score 85;  DB 10;  Length 395;
  Best Local Similarity  33.3%;  Pred. No. 8.31e+00;
  Matches    14;  Conservative   12;  Mismatches  15;  Indels   1;  Gaps   1;

Db     27 TLLALGKAFPSQVVPQENLVEGFLRDTKCDDAFIKEKLEHLC 68
          |||: : :  : :  | :| |:|| ::  : :|: | || ||
Qy     12 TLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFL-ELLEKLC 52


RESULT   13
ID   O42453      PRELIMINARY;      PRT;   448 AA.
AC   O42453;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SERPIN PRECURSOR.
OS   PETROMYZON MARINUS (SEA LAMPREY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; CEPHALASPIDOMORPHI;
OC   PETROMYZONTIFORMES; PETROMYZONTIDAE; PETROMYZON.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LARVAL LIVER;
RA   ROBSON P., YOUSON J.H., KEELEY F.W.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF009964; G2275601; -.
DR   PFAM; PF00079; serpin; 1.
KW   SIGNAL.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    448       SERPIN.
SQ   SEQUENCE   448 AA;  49044 MW;  6E233618 CRC32;

  Query Match           15.2%;  Score 85;  DB 13;  Length 448;
  Best Local Similarity  34.9%;  Pred. No. 8.31e+00;
  Matches    15;  Conservative   12;  Mismatches  14;  Indels   2;  Gaps   2;

Db     82 LMMLLAGSGDKTETQLTNALRL-Q-FLRDPNPQASFQALVSKL 122
          |:: | ||:  |   :|  |:| : || ::: ::||  |: ||
Qy      9 LLLTLLGSSHGTGPGMTLQLKLKESFLTNSSYESSFLELLEKL 51
```

```
RESULT   14
ID   O45649      PRELIMINARY;      PRT;   676 AA.
AC   O45649;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   K03D7.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MATTHEWS L.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z81562; E1188135; -.
SQ   SEQUENCE   676 AA;  77673 MW;  366C5777 CRC32;

  Query Match           15.2%;  Score 85;  DB 5;  Length 676;
  Best Local Similarity  41.5%;  Pred. No. 8.31e+00;
  Matches    17;  Conservative    9;  Mismatches  13;  Indels   2;  Gaps   2;

Db    380 LELKEINLTNAKEFKGESFQLSVPMELLLHF-SNVSVTLQQ 419
          | |||  |||:  : :  ::|   : ||||: |  ||.||::
Qy     28 LKLKESFLTNSS-YESSFLELLEKLCLLLHLPSGTSVTLHH 67


RESULT   15
ID   O43190      PRELIMINARY;      PRT;   371 AA.
AC   O43190;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE   PURINERGIC P2Y11 RECEPTOR.
GN   P2Y11.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   COMMUNI D., GOVAERTS C., PARMENTIER M., BOEYNAEMS J.M.;
RL   J. BIOL. CHEM. 0:0-0(1997).
DR   EMBL; AF030335; G2674120; -.
SQ   SEQUENCE   371 AA;  40174 MW;  5686F41C CRC32;

  Query Match           15.1%;  Score 84;  DB 4;  Length 371;
  Best Local Similarity  38.1%;  Pred. No. 1.09e+01;
  Matches    16;  Conservative   12;  Mismatches  11;  Indels   3;  Gaps   2;

Db    209 LGCGLPLLLTLAAYGALGRAVLRSPGMTVAEKLRVAALVASG 250
          :| ||||:| |:  |: :::   :||||:  ||: : |  |:
Qy      1 MGSGLPLVLLLTLLGS-SHGT--GPGMTLQLKLKESFLTNSS 39


Search completed: Fri Oct 22 18:44:57 1999
Job time : 51 secs.
```

****

MPSRCH (TM)

****

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 16:51:21 1999; MasPar time 456.53 Seconds
1390.417 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-2
Description: (1-229) from US09092296.seq
Perfect Score: 229
N.A. Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCCTTCAGGGACCA 229
Comp: TGGCCCTGAAGTCACAGAGG.....GGTAGAGGGAAGTCCCTGGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl58
1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi
Database: genbank111
17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 9.569; Variance 4.802; scale 1.993

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	108	47.2	47323	31	AC005937	Homo sapiens clone UWG	8.69e-62
	2	46	20.1	7218	25	I66494	Sequence 14 from paten	9.04e-15
c	3	33	14.4	215	25	I28278	Sequence 5 from patent	4.40e-06
	4	33	14.4	216021	31	HUAC004787	Homo sapiens Chromosom	4.40e-06
	5	31	13.5	965	25	AR024229	Sequence 22 from paten	7.82e-05
	6	29	12.7	74371	31	AC005369	Homo sapiens chromosom	1.30e-03
c	7	29	12.7	216021	31	HUAC004787	Homo sapiens Chromosom	1.30e-03
	8	27	11.8	215	25	I28278	Sequence 5 from patent	1.98e-02
c	9	27	11.8	965	25	AR024229	Sequence 22 from paten	1.98e-02
	10	26	11.4	1056	23	MVU87256	Mustela vison GT dinuc	7.48e-02
	11	25	10.9	565	25	E04076	gDNA encoding envelope	2.76e-01
	12	25	10.9	60966	31	AC003030	Homo sapiens chromosom	2.76e-01
c	13	24	10.5	60	25	A62989	Sequence 1 from Patent	9.88e-01
c	14	24	10.5	1056	23	MVU87256	Mustela vison GT dinuc	9.88e-01
	15	23	10.0	3290	29	HUMPFUCAS	H.sapiens fucosidase p	3.44e+00
c	16	23	10.0	133457	30	AC003999	Human PAC clone DJ1139	3.44e+00
	17	23	10.0	203418	19	AC004947	Homo sapiens clone DJ1	3.44e+00
c	18	22	9.6	30	25	A62994	Sequence 6 from Patent	1.16e+01
c	19	22	9.6	108	21	D87227	Trypanosoma cruzi mRNA	1.16e+01
	20	22	9.6	288	29	HS156G5R	H.sapiens CpG island D	1.16e+01
c	21	22	9.6	1738	29	HUMGOLGINB	Human (clone SY2/10) g	1.16e+01
c	22	22	9.6	2180	21	DROSYT	D.melanogaster synapto	1.16e+01
	23	22	9.6	2273	32	MMBAKEXN	Mus musculus Bak gene,	1.16e+01
c	24	22	9.6	3088	11	E10775	DNA encoding part of A	1.16e+01
c	25	22	9.6	3088	25	AR014574	Sequence 1 from patent	1.16e+01
c	26	22	9.6	3890	30	HSA010901	Homo sapiens MUC4 gene	1.16e+01
	27	22	9.6	4472	17	NMSERBORF	Neisseria meningitidis	1.16e+01
c	28	22	9.6	6640	29	D63997	Homo sapiens mRNA for	1.16e+01
	29	22	9.6	67919	20	AC006542	WORKING DRAFT SEQUENCE	1.16e+01
c	30	22	9.6	79482	19	AP000029	Homo sapiens genomic D	1.16e+01
	31	22	9.6	82098	31	AC006252	Homo sapiens 3p21.1 co	1.16e+01
	32	22	9.6	82201	19	HS593C16	Human DNA sequence ***	1.16e+01
c	33	22	9.6	124700	22	AC005558	Drosophila melanogaste	1.16e+01
	34	22	9.6	135039	31	AC006060	Homo sapiens 3p22-8 PA	1.16e+01
	35	22	9.6	151187	19	HS460J8	Human DNA sequence ***	1.16e+01
c	36	22	9.6	175339	31	AC005772	Homo sapiens chromosom	1.16e+01
c	37	22	9.6	175793	31	AC005920	Homo sapiens chromosom	1.16e+01
	38	22	9.6	192853	19	AC005096	Homo sapiens clone RG3	1.16e+01
c	39	22	9.6	241407	20	AC003059	Mouse Chromosome 10 BA	1.16e+01
	40	21	9.2	66	25	I41364	Sequence 144 from pate	3.79e+01
c	41	21	9.2	68	25	I41363	Sequence 143 from pate	3.79e+01
	42	21	9.2	2976	21	PFAPA2B	Plasmodium knowlesi P-	3.79e+01
	43	21	9.2	3043	32	AB010141S4	Mus musculus gene for	3.79e+01
	44	21	9.2	130560	19	HS508I15	Human DNA sequence ***	3.79e+01
c	45	21	9.2	149298	19	AP000031	Homo sapiens genomic D	3.79e+01

ALIGNMENTS

RESULT 1
LOCUS AC005937 47323 bp DNA PRI 05-NOV-1998
DEFINITION Homo sapiens clone UWGC:370M23.002 from 6p21, complete sequence.
ACCESSION AC005937
NID g3845393
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 47323)
AUTHORS Janer,M., Guillaudeux,T., Vu,Q., Kutyavin,T., Harter,H. and Geraghty,D.E.
TITLE Large scale sequence analysis of the human MHC class I region
JOURNAL Unpublished (1998)
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
REFERENCE 2 (bases 1 to 47323)
AUTHORS Geraghty,D.E. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhcrc.org)
COMMENT Overlapping Sequences:
5': UWGC:370m23.013 (Genbank Accession: AC005530)
3': UWGC:y67c112 (Genbank Accession: AC004211)

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than

1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 75.5%
DS or two chemistry coverage: 98.9%
Single stranded regions: 3

Sequence Validation:
This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BglII		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
1069.11	1050.00	889.55	866.00	30541.40	30653.00
20320.67	20855.00	1050.18	1015.00	3279.08	3231.00
2171.50	2147.00	7268.78	7196.00		
2560.20	2531.00	10085.80	9992.00		
4335.42	4269.00	11212.78	11131.00		
2698.62	2628.00				
1927.50	1887.00				
3130.46	3090.00				
2166.69	2129.00				
2044.67	2005.00				

```
FEATURES             Location/Qualifiers
     source          1. .47323
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21"
                     /sub_clone="UWGC:370M23.002"
                     /clone_lib="Research Genetics BAC Library"
     repeat_region   3647. .3932
                     /rpt_family="Alu"
     repeat_region   complement(4999. .5277)
                     /rpt_family="Alu"
     repeat_region   6285. .6572
                     /rpt_family="Alu"
     repeat_region   complement(6972. .7050)
                     /rpt_family="MLT1"
     repeat_region   7286. .7584
                     /rpt_family="Alu"
     repeat_region   complement(8164. .8609)
                     /rpt_family="Alu"
     repeat_region   complement(21287. .21895)
                     /rpt_family="Alu"
     repeat_region   22715. .22957
                     /rpt_family="Alu"
     repeat_region   25510. .25802
                     /rpt_family="Alu"
     repeat_region   27835. .28010
                     /rpt_family="MER20"
     repeat_region   31295. .31594
                     /rpt_family="Alu"
     repeat_region   33515. .33767
                     /rpt_family="Alu"
     repeat_region   34223. .34290
                     /rpt_family="MIR"
     repeat_region   37372. .37648
                     /rpt_family="Alu"
     repeat_region   38526. .38700
                     /rpt_family="MER3"
     repeat_region   39583. .40010
                     /rpt_family="Alu"
     repeat_region   40046. .40156
                     /rpt_family="Alu"
     repeat_region   43194. .43372
                     /rpt_family="MER5"
     variation       43325
                     /note="clonal variation with 3' overlapping clone"
     variation       44149
                     /note="clonal variation with 3' overlapping clone"
     variation       44451
                     /note="clonal variation with 3' overlapping clone"
     variation       44537
                     /note="clonal variation with 3' overlapping clone"
     variation       44814
                     /note="clonal variation with 3' overlapping clone"
     variation       44965
                     /note="clonal variation with 3' overlapping clone"
     variation       45760
                     /note="clonal variation with 3' overlapping clone"
     variation       45900
                     /note="clonal variation with 3' overlapping clone"
     variation       46851
                     /note="clonal variation with 3' overlapping clone"
     variation       46859
                     /note="clonal variation with 3' overlapping clone"
     variation       47032
                     /note="clonal variation with 3' overlapping clone"
     variation       47240. .47256
                     /note="clonal variation with 3' overlapping clone -
                     insertion of 17bp repeat"
BASE COUNT    11556 a  11489 c  12284 g  11994 t
ORIGIN

  Query Match           47.2%;  Score 108;  DB 31;  Length 47323;
  Best Local Similarity 100.0%;  Pred. No. 8.69e-62;
  Matches   108;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  35121 ACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCT 35180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 ACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCT 60

Db  35181 GCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 35228
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 108


RESULT    2
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

```
  Query Match           20.1%;  Score 46;  DB 25;  Length 7218;
  Best Local Similarity   2.3%;  Pred. No. 9.04e-15;
  Matches     5;  Conservative  129;  Mismatches  88;  Indels   0;  Gaps   0;

Db   1219 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1278
          ::::  : :::::::: ::::     : :  :  :: :: :    :::   :: ::::::
Qy      8 CTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTT 67

Db   1279 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1338
          ::::::::::  ::::::::  :  ::: : :    :    ::   : :  :::: :
Qy     68 GTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAA 127

Db   1339 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1398
          ::     ::       :::::::: :   :::::::: :   :::  ::::::    ::
Qy    128 CTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTG 187

Db   1399 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440
          :::      ::: :::::::::: :::::::::  | ||||
Qy    188 CTTGAAAAGTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCA 229


RESULT    3
LOCUS       I28278        215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
VERSION     I28278.1  GI:1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="unknown"
BASE COUNT       15 a       8 c      25 g      26 t     141 others
ORIGIN

  Query Match           14.4%;  Score 33;  DB 25;  Length 215;
  Best Local Similarity  14.6%;  Pred. No. 4.40e-06;
  Matches    26;  Conservative   73;  Mismatches  77;  Indels   2;  Gaps   2;

Db      1 MTNVTMSSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNK 60
          :| : :::   :: : |:  ::|:::|     : |    : :|:|:  ||:::  :: : :
Cp    185 ATTCCAGGAAGCTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTC-AGT 127

Db     61 YSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTD-GNRSGADS 119
          :::|   :    :   : ::: :| :   ::  :|::: |:| :::    : | ::||
Cp    126 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 67

Db    120 YGSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTKKHAKNSADGKVGSKNNGDR 177
          |::: |  :  :       |:::: :   ::|::  |:: |: :|:| : :   |::
Cp     66 AGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATGGAGGAGACACTGAA 9


RESULT    4
LOCUS       HUAC004787 216021 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
NID         g3337381
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.  The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .216021
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="#16q21-22"
                     /clone="A-952F10"
     STS             27765. .27872
                     /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02280"
     STS             73826. .73943
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     STS             175801. .175945
                     /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09703"
     STS             175810. .175945
                     /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09935"
     STS             199463. .199572
                     /note="9824, WI-3555, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G04338"
BASE COUNT    60960 a   51778 c   49172 g   53987 t    124 others
ORIGIN

  Query Match           14.4%;  Score 33;  DB 31;  Length 216021;
  Best Local Similarity  12.2%;  Pred. No. 4.40e-06;
  Matches    18;  Conservative   72;  Mismatches  57;  Indels   0;  Gaps   0;

Db   1436 WYCSSYTYCCYYCSRARRRGKKKYMWKSMRRARRRGSAGKKKKKYYYYYCYYYYYYYYCM 1495
          ::| :: :||:  ::  :  ::: :  : :  ::: : |:: :::::::  ::::::::|
Qy     18 TCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCT 77

Db   1496 GRAMMAAAWYYKRRSCMAWYYYMRGRARKTYYYMARRRGCARSYKKYMAMYWAAWTTWCA 1555
           :|::    ::::   :  :  :|:|     :: ::    ::::: :|::: |:  : |
Qy     78 TGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTGA 137

Db   1556 AAAMAAMYTTYYWMMWYYKKWMYTCCT 1582
          |  :  ::||::: :    ::::|||
Qy    138 AGGAGTCTTTTCTGACAAATTCCTCCT 164


RESULT    5
```

```
LOCUS       AR024229      965 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 22 from patent US 5795961.
ACCESSION   AR024229
NID         g3977523
VERSION     AR024229.1  GI:3977523
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and
            Kitamura,K.
  TITLE     Recombinant human anti-Lewis b antibodies
  JOURNAL   Patent: US 5795961-A 22 18-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .965
                     /organism="unknown"
BASE COUNT      192 a    170 c    226 g    205 t    172 others
ORIGIN

  Query Match           13.5%;  Score 31;  DB 25;  Length 965;
  Best Local Similarity  20.2%;  Pred. No. 7.82e-05;
  Matches    24;  Conservative   52;  Mismatches  42;  Indels   1;  Gaps   1;

Db    845 SVTAADTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHTVKDMTSSSSASVGDRVTTCRSST 904
          ::|  :| ::  ::| :::::|| ::   |: :: :|::::|::::   |: :| | : |
Qy     19 CCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTT 78

Db    905 THGNGNTYYWYKGKAKYRVSNRSGVSRSGSGSGTDYTTSSDATYYCGTHARTGGTKVKG 963
           :    |:::::| |:: :  : |: :: :|:  :  |  :| :: | :| ||  :::|
Qy     79 GACC-CTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG 136


RESULT    6
LOCUS       AC005369    74371 bp    DNA             PRI       01-AUG-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 119j3 (LBNL H175), complete
            sequence.
ACCESSION   AC005369
NID         g3367505
VERSION     AC005369.1  GI:3367505
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 74371)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 74371)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 74371)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .74371
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="119j3"
                     /chromosome="5"
                     /note="LBNL H175"
     repeat_region   893. .1030
                     /rpt_family="Alu"
     repeat_region   2295. .2438
                     /rpt_family="Alu"
     repeat_region   2818. .2859
                     /note="(GT)21"
                     /rpt_type=tandem
                     /rpt_unit=GT
     misc_feature    join(3246. .3410,3721. .3828)
                     /standard_name="RLF"
                     /note="65% & 69% protein identity GenPept:U22377"
     repeat_region   3431. .3724
                     /rpt_family="Alu"
     repeat_region   3707. .3728
                     /note="(A)22"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   4366. .4661
                     /rpt_family="Alu"
     repeat_region   5327. .5602
                     /rpt_family="Alu"
     repeat_region   6586. .6956
                     /rpt_family="L1"
     repeat_region   6647. .6684
                     /note="(CA)19"
                     /rpt_type=tandem
                     /rpt_unit=CA
     repeat_region   7113. .7373
                     /rpt_family="Alu"
     repeat_region   complement(7830. .8185)
                     /standard_name="possible repeat"
     repeat_region   8258. .8503
                     /rpt_family="Alu"
     repeat_region   9070. .9387
                     /rpt_family="Alu"
     repeat_region   complement(9740. .9845)
                     /rpt_family="MER42"
     repeat_region   complement(10440. .11015)
                     /rpt_family="Alu"
     repeat_region   11950. .12250
                     /rpt_family="Alu"
     repeat_region   12057. .12085
                     /note="(A)29"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   12365. .12645
                     /rpt_family="Alu"
     repeat_region   13727. .13750
                     /note="(AC)12"
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   13783. .14024
                     /rpt_family="L1"
     repeat_region   14175. .14470
                     /rpt_family="Alu"
     repeat_region   complement(14906. .15259)
                     /standard_name="possible repeat"
     repeat_region   15300. .15613
                     /rpt_family="Alu"
     repeat_region   16671. .16690
                     /note="(T)20"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(16675. .16977)
                     /rpt_family="Alu"
     repeat_region   complement(16993. .17085)
                     /rpt_family="MER42"
     repeat_region   complement(17678. .18276)
                     /rpt_family="Alu"
     repeat_region   19305. .19583
                     /rpt_family="Alu"
     repeat_region   19914. .19945
```

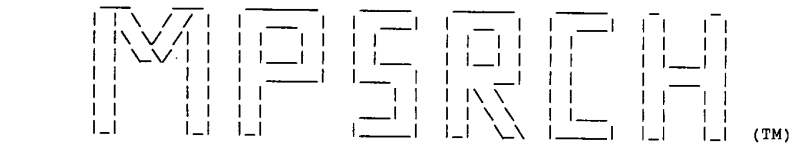
```
                /note="(GTTT)8"
                /rpt_type=tandem
                /rpt_unit=GTTT
repeat_region   complement(19943. .20222)
                /rpt_family="Alu"
repeat_region   21202. .21496
                /rpt_family="Alu"
repeat_region   21736. .22035
                /rpt_family="Alu"
repeat_region   22017. .22038
                /note="(A)22"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   22321. .22457
                /rpt_family="Alu"
repeat_region   complement(22638. .22981)
                /standard_name="possible repeat"
repeat_region   23022. .23326
                /rpt_family="Alu"
repeat_region   23473. .23761
                /rpt_family="Alu"
repeat_region   23744. .23767
                /note="(A)24"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   complement(24639. .24694)
                /rpt_family="MER42"
repeat_region   complement(25349. .25713)
                /rpt_family="Alu"
repeat_region   complement(25727. .26471)
                /rpt_family="Alu"
repeat_region   27191. .27477
                /rpt_family="Alu"
repeat_region   27774. .28057
                /rpt_family="Alu"
repeat_region   28040. .28066
                /note="(A)27"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   28712. .28930
                /rpt_family="MER20"
misc_feature    complement(28769. .28838)
                /note="GRAIL 2 excellent exon, frame 0"
repeat_region   28987. .29214
                /rpt_family="Alu"
misc_feature    join(29485. .29559,29598. .29779)
                /note="95% identity dbSTS:G14522 (SHGC-11312)"
misc_feature    29495. .29976
                /standard_name="AI027942"
                /note="100% identity EST ov84a10.x1"
misc_feature    complement(30401. .30536)
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   complement(30682. .30733)
                /rpt_family="MIR"
misc_feature    complement(31573. .31724)
                /note="GRAIL 2 excellent exon, frame 1"
misc_feature    complement(32159. .32232)
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   32388. .32488
                /rpt_family="MLT1"
repeat_region   32617. .32908
                /rpt_family="Alu"
repeat_region   32977. .33088
                /rpt_family="MLT1"
repeat_region   complement(33670. .33785)
                /rpt_family="Alu"
misc_feature    complement(34021. .34144)
                /note="GRAIL 2 excellent exon, frame 2"
misc_feature    complement(35238. .35331)
                /note="GRAIL 2 excellent exon, frame 1"
repeat_region   36392. .36663
                /rpt_family="Alu"
misc_feature    36901. .37222
                /note="100% identity EST ou55c09.x1"
                /db_xref="dbEST:AI025011"
STS             36901. .37164
                /standard_name="STSG-9983"
                /db_xref="dbSTS:G26554"
misc_feature    complement(37595. .37654)
                /note="GRAIL 2 excellent exon, frame 0"
misc_feature    complement(join(38063. .38218,38462. .38578,38741. .38995,
                39071. .39205,39532. .39630,39935. .40048,40300. .40410,
                40503. .40661,41868. .41972,42103. .42225,42492. .42569,
                44379. .44507))
                /standard_name="histidyl-tRNA synthetase"
                /note="78%-100% protein identity GenPept:U18937"
misc_feature    complement(38069. .38215)
                /note="GRAIL 2 excellent exon, frame 0"
...
Note: remainder of annotations omitted.

  Query Match          12.7%;  Score 29;  DB 31;  Length 74371;
  Best Local Similarity  16.7%;  Pred. No. 1.30e-03;
  Matches    14;  Conservative    45;  Mismatches  24;  Indels    1;  Gaps    1;

Db  15929 CMSRSKSKRGWGYRSWKKYRCAMWMTCKSS-KCWCWSYRMRMKCYSCSYCYCSSGKKYWC 15987
          |:::: ::  :|::  : : | ::: | :: :| |::::  : |:  :    : |::::
Qy      2 CCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTG 61

Db  15988 RCSMYWYTCYYSYKYYWSMSYCTC 16011
           |:::: ||:::::::::::::|||
Qy     62 CCCCTTGTCCTCCTCTTGACCCTC 85


RESULT    7
LOCUS       HUAC004787 216021 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
NID         g3337381
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
```

```
FEATURES             Location/Qualifiers
     source          1. .216021
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="#16q21-22"
                     /clone="A-952F10"
     STS             27765. .27872
                     /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02280"
     STS             73826. .73943
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     STS             175801. .175945
                     /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09703"
     STS             175810. .175945
                     /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09935"
     STS             199463. .199572
                     /note="9824, WI-3555, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G04338"
BASE COUNT    60960 a  51778 c  49172 g  53987 t    124 others
ORIGIN

  Query Match           12.7%;  Score 29;  DB 31;  Length 216021;
  Best Local Similarity  14.7%;  Pred. No. 1.30e-03;
  Matches    19;  Conservative   60;  Mismatches  50;  Indels   0;  Gaps   0;

Db   1457 KKYMWKSMRRARRRGSAGKKKKKYYYYYCYYYYYYYYCMGRAMMAAAWYYKRRSCMAWYY 1516
           :: : ::::: :::|:||: ::    :: :::   :      :  || :::::: :|:
Cp    222 TGAAGGGAGATGGAGGAGGAGGCAGACTTTTCAAGCAATTCCAGGAAGCTGGACTCATAG 163

Db   1517 YMRGRARKTYYYYMARRRGCARSYKKYMAMYWAAWTTWCAAAAMAAMYTTYYWMMWYYKKW 1576
           ::|:| :| ::: :::  |  : :::| ::  : |:  |||   ::  ::  : :::::
Cp    162 GAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTT 103

Db   1577 MYTCCTGAG 1585
           ::   ||||
Cp    102 CCATGTGAG 94


RESULT    8
LOCUS       I28278       215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
VERSION     I28278.1  GI:1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="unknown"
BASE COUNT       15 a      8 c     25 g     26 t    141 others
ORIGIN

  Query Match           11.8%;  Score 27;  DB 25;  Length 215;
  Best Local Similarity  20.0%;  Pred. No. 1.98e-02;
  Matches    22;  Conservative   42;  Mismatches  45;  Indels   1;  Gaps   1;

Db     18 CNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAK 77
          | :: : :| |  ::| :   :  |:|: :| :::  : |:  : :|   :|   :|| :
Qy     85 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG-AAGGAGT 143

Db     78 THYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
             :::|:| :: |: : :| ::  :   :: | |:   :||: :|: : |:
Qy    144 CTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAA 193


RESULT    9
LOCUS       AR024229      965 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 22 from patent US 5795961.
ACCESSION   AR024229
NID         g3977523
VERSION     AR024229.1  GI:3977523
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and
            Kitamura,K.
  TITLE     Recombinant human anti-Lewis b antibodies
  JOURNAL   Patent: US 5795961-A 22 18-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .965
                     /organism="unknown"
BASE COUNT      192 a    170 c    226 g    205 t    172 others
ORIGIN

  Query Match           11.8%;  Score 27;  DB 25;  Length 965;
  Best Local Similarity  15.6%;  Pred. No. 1.98e-02;
  Matches    22;  Conservative   59;  Mismatches  60;  Indels   0;  Gaps   0;

Db    790 GGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKRGRVTMADTSSNSRSSVTAA 849
          ||::: :| | ::     :  ::|: :| :|| :|    :: :: : :|:    :  :  |
Cp    202 GGCAGACTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGGAGGAATTTGTCAGAAAAGA 143

Db    850 DTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHTVKDMTSSSSASVGDRVTTCRSSTTHGNG 909
            | :::|::   :::: |  : || :|:::: :  :|:  :   |:::   :::   | |
Cp    142 CTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAG 83

Db    910 NTYYWYKGKAKYRVSNRSGVS 930
            ::: : :   ::: ::|::
Cp     82 GGTCAAGAGGAGGACAAGGGG 62


RESULT   10
LOCUS       MVU87256     1056 bp    DNA             MAM       02-JAN-1999
DEFINITION  Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION   U87256
NID         g4099442
VERSION     U87256.1  GI:4099442
KEYWORDS    .
SOURCE      American mink.
  ORGANISM  Mustela vison
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
            Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /chromosome="1"
                     /map="1q"
                     /note="primers: 1167F: agcccctgcatatctacttctt, 1167R:
                     gaggatcttaccgctgttgag"
     primer_bind     98. .119
                     /standard_name="1167F"
     primer_bind     complement(300. .320)
                     /standard_name="1167R"
BASE COUNT      211 a    221 c    210 g    225 t    189 others
ORIGIN
```

```
  Query Match           11.4%;  Score 26;  DB 23;  Length 1056;
  Best Local Similarity  12.0%;  Pred. No. 7.48e-02;
  Matches    11;  Conservative   48;  Mismatches  33;  Indels   0;  Gaps   0;

Db    578 RHDVBMWSKWCWGKKSCCKSTGKDKMSGCAYCGKRMRRCSRYMMMRMCVGSGTWARCCCC 637
          :: :: ::::|:|:: |  :|  ::::     | :::: :::::: : :|:||  :  |
Qy      1 ACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCT 60

Db    638 DKKSGSGDKSHCKSKRKYKWMDRYHBCKSMCA 669
          :  :   ::::  :: :::::  ::: :: ||
Qy     61 GCCCCTTGTCCTCCTCTTGACCCTCCTTGGCA 92


RESULT   11
LOCUS       E04076        565 bp    RNA             PAT       29-SEP-1997
DEFINITION  gDNA encoding envelope region of type C hepatitis virus.
ACCESSION   E04076
NID         g2172286
VERSION     E04076.1  GI:2172286
KEYWORDS    JP 1992349885-A/1.
SOURCE      Hepatitis C virus.
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepatitis C-like viruses.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Morinaga,T., Chayama,K., Kumada,H. and Ichikawa,Y. .
  TITLE     NUCLEIC ACID FRAGMENT WITHIN ENVELOPE REGION OF HEPATITIS C VIRUS
            AND METHOD FOR DETECTING THE SAME
  JOURNAL   Patent: JP 1992349885-A 1 04-DEC-1992;
            TEIJIN LTD
COMMENT     OS   Hepatitis C virus
            PN   JP 1992349885-A/1
            PD   04-DEC-1992
            PF   29-MAY-1991 JP 1991152169
            PI   MORINAGA TSUTAE, CHAYAMA KAZUAKI, KUMADA HIROMITSU, PI
            ICHIKAWA YATARO
            PC   C12N15/10,C12Q1/68,C12Q1/70//C12N15/11;
            CC   strandedness: Single;
            CC   topology: Linear;.
FEATURES            Location/Qualifiers
     source         1. .565
                    /organism="Hepatitis C virus"
                    /db_xref="taxon:11103"
BASE COUNT       60 a      93 c     107 g      85 t     220 others
ORIGIN

  Query Match           10.9%;  Score 25;  DB 25;  Length 565;
  Best Local Similarity  34.1%;  Pred. No. 2.76e-01;
  Matches    30;  Conservative   29;  Mismatches  28;  Indels   1;  Gaps   1;

Db    400 GGRTCYGTYTTYCTYRTYTCYCABYTGTTTACCYTCTMBMCYMGSMDGYAYGDGACRDYR 459
          ||:||:|   | | : |:  :| ::| || |||:||  :  ::|:   :|:|: ||:: :
Qy     50 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 109

Db    460 CARGRYTGYAAYTGYTCDMTYTAYSCYG 487
          | :|::  : |:| : |:: : | :|:|
Qy    110 CCGGGTA-TGACTTTGCAACTGAAGCTG 136


RESULT   12
LOCUS       AC003030     60966 bp    DNA             PRI       06-JAN-1999
DEFINITION  Homo sapiens chromosome 19, overlapping cosmids R29828 and F25496,
            complete sequence.
ACCESSION   AC003030
NID         g4092821
VERSION     AC003030.1  GI:4092821
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
            Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S.,
            Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 2 Mb contig in 19p12 between UBA52 and
            D19S455
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1997) Human Genome Center, Joint Genome
            Institute/ Lawrence Livermore National Laboratory, 7000 East Ave.,
            Livermore, CA 94551, USA
REFERENCE   3  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE   4  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     On Jan 4, 1999 this sequence version replaced gi:2570224.
            Map and sequence oriented from p telomere to centromere. This
            accession is comprised of overlapping cosmids R29828 (bases 1 to
            40,974) and F25496 (bases 23,336 to 60,966). R29828 is separated
            from cosmid F23858 (AC004475) to the left by a sequence gap of
            approximately 14 kb, which is to be filled by sequencing a
            restriction fragment from cosmid R27236 (currently in progress).
            Cosmid F25496 is separated from cosmid R31863 to the right by an
            expected sequence gap of at least 10 kb. Additional chr 19 map and
            sequence information are available at:
            http://www-bio.llnl.gov/genome/genome.html.
FEATURES            Location/Qualifiers
     source         1. .60966
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="R29828-F25496"
                    /chromosome="19"
                    /map="19p12 between UBA52 and D19S455."
                    /cell_line="UV5HL9-5B for F25496, and 5HL2-B for R29828"
                    /clone_lib="LL19NC02 (for F25496) and LL19NC03 (for
                    R29828) chromosome 19-specific cosmid libraries"
                    /note="Cosmid libraries constructed at LLNL from
                    flow-sorted chromosomes from human-hamster hybrid cell
                    lines UV5HL9-5B and 5HL2-B, each of which carries
                    chromosome 19 as its only human chromosome."
     repeat_region  18. .219
                    /rpt_family="AluJo"
     repeat_region  530. .824
                    /rpt_family="AluY"
     repeat_region  930. .1066
                    /rpt_family="FLAM_C"
     misc_feature   complement(1148. .1315)
                    /note="predicted exon, program: grail2exons_human_1.3,
                    frame: 2, quality: good, score: 51.000"
     repeat_region  complement(1477. .1706)
                    /rpt_family="AluJo"
     repeat_region  2105. .2406
                    /rpt_family="AluSp"
     repeat_region  2513. .2809
                    /rpt_family="AluJo"
     misc_feature   5028. .5045
                    /note="DDS similarity to overlapping ESTs:~AA963316
                    UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                    clone UI-R-E1- gi-c-06-0-UI 3'; (220. .237); 100%
                    identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                    Soares Rattus sp. cDNA clone RKIBE38 3' end; (240. .257);
```

```
                100% identity.~"
CDS             join(<5028. .5045,5432. .5497,6480. .6575,8155. .8249,
                10156. .10183,10562. .10717,11845. .11991,12086. .12176,
                13147. .13250,14614. .14679,18199. .18248,18465. .18542,
                18635. .18746,23665. .23755,24574. .24709,37283. .37493)
                /note="Hypothetical human protein (partial)"
                /codon_start=1
                /evidence=not_experimental
                /product="R29828_1"
                /protein_id="AAD03161.1"
                /db_xref="PID:g4106983"
                /db_xref="GI:4106983"
                /translation="WLISQQIPQFEDVKFEAASLLSELYCQENSVDAAKPLLRKAIQI
                SQQTPYWHCRLLFQLAQLHTLEKDLVSACDLLGVGAEYARVVGSEYTRALFLLSKGML
                LLMERKLQEVHPLLTLCGQIVENWQGNPIQKESLRVFFLVLQVTHYLDAGQVKSVKPC
                LKQLQQCIQTISTLHDDEILPSNPADLFHWLPKEHMCVLVYLVTVMHSMQAGYLEKAQ
                KYTDKALMQLEKLKMLDCSPILSSFQVILLEHIIMCRLVTGHKATALQEGLYCVSVNC
                MDNAEAQFTTALRLYSLLERINPDHSFPVSSHCLRAAAFYVRGLFSFFQGRYNEAKRF
                LRETLKMSNAEDLNRLTACSLVLLGHIFYVLGNHRESNNMVVPAMQLASKIPDMSVQL
                WSSALLRDLNKACGNAMDAHEAAQMHQNFSQQLLQDHIEACSLPEHNLITVRVGHWEG
                IAATRLPVAFMEGQSTTTLYSADIWKHIQEDTEGSESAWAPVFRCEIKSLRPQTRAII
                LLPPLD"
misc_feature    complement(5186. .5264)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 86.000"
misc_feature    5432. .5497
                /note="DDS similarity to overlapping ESts:~AA963316
                UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                clone UI-R-E1- gi-c-06-0-UI 3'; (238. .303); 88%
                identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                Soares Rattus sp. cDNA clone RKIBE38 3' end; (258. .323);
                88% identity."
repeat_region   complement(5602. .5883)
                /rpt_family="AluJo"
misc_feature    6480. .6575
                /note="DDS similarity to overlapping ESTs:~AA963316
                UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                clone UI-R-E1- gi-c-06-0-UI 3'; (304. .399); 89%
                identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                Soares Rattus sp. cDNA clone RKIBE38 3' end; (324. .419);
                89% identity.~~(6480. .6660) predicted exon, program:
                grail2exons_human_1.3, frame: 2, quality: excellent,
                score: 76.000"
misc_feature    8155. .8249
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 94.000~~DDS
                similarity to overlapping ESTs:~(8155. .8237) AA963316
                UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                clone UI-R-E1- gi-c-06-0-UI 3'; (400. .482); 88%
                identity.~~(8155. .8240) AA893275 EST197078 Normalized rat
                kidney, Bento Soares Rattus sp. cDNA clone RKIBE38 3' end;
                (420. .505); 85% identity."
repeat_region   8351. .8477
                /rpt_family="AluJb"
repeat_region   8480. .8768
                /rpt_family="AluSc"
misc_feature    8778. .8837
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 83.000"
repeat_region   complement(8872. .8996)
                /rpt_family="AluJb"
repeat_region   9052. .9196
                /rpt_family="AluSx"
repeat_region   9197. .9493
                /rpt_family="AluY"
repeat_region   9494. .9658
                /rpt_family="AluSx"
repeat_region   9668. .9967
                /rpt_family="AluY"
misc_feature    10156. .10183
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 86.000"
misc_feature    10562. .10721
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 100.000"
misc_feature    11845. .11991
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 97.000"
misc_feature    12086. .12176
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 96.000"
misc_feature    13147. .13250
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 100.000"
repeat_region   13440. .13741
                /rpt_family="AluSx"
misc_feature    14614. .14679
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 100.000"
misc_feature    16463. .16747
                /note="DDS similarity to AA212842 mw82d12.r1 Soares mouse
                NML Mus musculus cDNA clone 677207 5'; (200. .468); 80%
                identity."
repeat_region   17395. .17687
                /rpt_family="AluJb"
misc_feature    18199. .18248
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 89.000"
misc_feature    18465. .18542
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 100.000"
misc_feature    18635. .18746
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 97.000"
repeat_region   complement(19286. .19309)
                /rpt_family="POLY_A"
repeat_region   19323. .19623
                /rpt_family="AluSx"
repeat_region   19632. .19928
                /rpt_family="AluSx"
repeat_region   19929. .20226
                /rpt_family="AluY"
repeat_region   20234. .20341
                /rpt_family="AluJo/FLAM"
repeat_region   complement(20344. .20655)
...
Note: remainder of annotations omitted.

  Query Match          10.9%;  Score 25;  DB 31;  Length 60966;
  Best Local Similarity  85.7%;  Pred. No. 2.76e-01;
  Matches    30;  Conservative    0;  Mismatches   5;  Indels   0;  Gaps   0;

Db  13976 TGTCCACGATGGGATCTGGGCTGCCCTTTGCCCTC 14010
          || |||| ||||| |||||||||||| ||| ||||
Qy     39 TGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTC 73


RESULT   13
LOCUS       A62989         60 bp    DNA              PAT       12-MAR-1998
DEFINITION  Sequence 1 from Patent WO9720068.
ACCESSION   A62989
NID         g3716861
VERSION     A62989.1  GI:3716861
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 60)
  AUTHORS   Oerum,H. and Seeger,C.
  TITLE     METHOD FOR GENERATING MULTIPLE DOUBLE STRANDED NUCLEIC ACIDS
  JOURNAL   Patent: WO 9720068-A 1 05-JUN-1997;
            BOEHRINGER MANNHEIM GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="unidentified"
                     /db_xref="taxon:32644"
```

```
BASE COUNT        0 a     30 c      0 g      0 t     30 others
ORIGIN

  Query Match           10.5%;  Score 24;  DB 25;  Length 60;
  Best Local Similarity   6.2%;  Pred. No. 9.88e-01;
  Matches     2;  Conservative   26;  Mismatches   4;  Indels   0;  Gaps   0;

Db      29 CCRRRRRRRRRRRRRRRRRRRRRRRRRRRR 60
           ||:::::::::  :::::::::: :::::: ::
Cp      90 CCAAGGAGGGTCAAGAGGAGGACAAGGGGCAG 59


RESULT   14
LOCUS       MVU87256     1056 bp    DNA             MAM       02-JAN-1999
DEFINITION  Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION   U87256
NID         g4099442
VERSION     U87256.1  GI:4099442
KEYWORDS    .
SOURCE      American mink.
  ORGANISM  Mustela vison
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
            Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /chromosome="1"
                     /map="1q"
                     /note="primers: 1167F: agcccctgcatatctacttctt, 1167R:
                     gaggatcttaccgctgttgag"
     primer_bind     98. .119
                     /standard_name="1167F"
     primer_bind     complement(300. .320)
                     /standard_name="1167R"
BASE COUNT      211 a    221 c    210 g    225 t    189 others
ORIGIN

  Query Match           10.5%;  Score 24;  DB 23;  Length 1056;
  Best Local Similarity  10.8%;  Pred. No. 9.88e-01;
  Matches     7;  Conservative   40;  Mismatches  17;  Indels   1;  Gaps   1;

Db     611 KRMRRCSRYMMMRMCVGSGTWARCCCCDKKSGSGDKSHCKSKRKYKWMDRYHBC-KSMCA 669
           ::::: : :: : :|:|:  :   || :::: : ::::  :::: ::::: ::  :::||
Cp     126 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 67

Db     670 MVRGK 674
           :::|:
Cp      66 AGGGG 62


RESULT   15
LOCUS       HUMPFUCAS    3290 bp    DNA             PRI       08-JAN-1995
DEFINITION  H.sapiens fucosidase pseudogene.
ACCESSION   M80816
NID         g189864
VERSION     M80816.1  GI:189864
KEYWORDS    alpha-L-fucosidase; lysosomal hydrolase; lysosome; pseudogene.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3290)
  AUTHORS   Seo,H.C., Willems,P.J., Kretz,K.A., Martin,B.M. and O'Brien,J.S.
  TITLE     Fucosidosis: four new mutations and a new polymorphism
  JOURNAL   Hum. Mol. Genet. 2 (4), 423-429 (1993)
  MEDLINE   93278392
FEATURES             Location/Qualifiers
     source          1. .3290
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2q31-q32"
     gene            735. .2761
                     /gene="FUCA1P"
     misc_feature    735. .2761
                     /gene="FUCA1P"
                     /note="homology to human a-L-fucosidase cDNA; G00-119-238;
                     putative"
BASE COUNT      930 a    660 c    681 g   1019 t
ORIGIN      chromosome 2.

  Query Match           10.0%;  Score 23;  DB 29;  Length 3290;
  Best Local Similarity  79.5%;  Pred. No. 3.44e+00;
  Matches    31;  Conservative    0;  Mismatches   8;  Indels   0;  Gaps   0;

Db    1596 TGCAACATGGCAATGTCTGACTTTGCAAGTGAATCTGAA 1634
           || |||| |||   || ||||||||||| |||| |||||
Qy     100 TGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTGAA 138


Search completed: Sun Oct 24 16:59:07 1999
Job time : 466 secs.
```



MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:07:21 1999; MasPar time 70.98 Seconds
691.329 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-2
Description: (1-229) from US09092296.seq
Perfect Score: 229
N.A. Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCCTTCAGGGACCA 229
Comp: TGGCCCTGAAGTCACAGAGG.....GGTAGAGGGAAGTCCCTGGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 7.766; Variance 4.929; scale 1.575

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	223	97.4	439	60	V84366	Human stomach carcino	8.25e-127
c	2	37	16.2	91	9	Q51746	Oligonucleotide probe	3.13e-08
	3	37	16.2	204	1	N81164	Base substituted E.co	3.13e-08
c	4	37	16.2	204	1	N81164	Base substituted E.co	3.13e-08
	5	36	15.7	91	9	Q51746	Oligonucleotide probe	1.12e-07
	6	31	13.5	114	12	Q70470	Generic DNA sequence	5.71e-05
c	7	31	13.5	114	12	Q70467	Generic DNA sequence	5.71e-05
	8	30	13.1	114	12	Q70468	Generic DNA sequence	1.93e-04
c	9	30	13.1	114	12	Q70470	Generic DNA sequence	1.93e-04
c	10	30	13.1	114	12	Q70468	Generic DNA sequence	1.93e-04
	11	29	12.7	114	12	Q70467	Generic DNA sequence	6.43e-04
	12	29	12.7	114	12	Q70465	Generic DNA sequence	6.43e-04
	13	29	12.7	114	12	Q70469	Generic DNA sequence	6.43e-04
c	14	28	12.2	114	12	Q70465	Generic DNA sequence	2.12e-03
	15	28	12.2	178	32	T76405	Human endothelin-1 an	2.12e-03
	16	27	11.8	114	12	Q70466	Generic DNA sequence	6.88e-03
	17	27	11.8	172	32	T76363	Human interleukin 8 a	6.88e-03
	18	26	11.4	114	12	Q70472	Generic DNA sequence	2.21e-02
c	19	26	11.4	114	12	Q70469	Generic DNA sequence	2.21e-02
c	20	25	10.9	39	7	Q51787	Mixed oligonucleotide	6.96e-02
c	21	25	10.9	114	12	Q70472	Generic DNA sequence	6.96e-02
	22	25	10.9	250	32	T76438	Substance P antisense	6.96e-02
c	23	24	10.5	114	12	Q70466	Generic DNA sequence	2.16e-01
c	24	24	10.5	114	12	Q70473	Generic DNA sequence	2.16e-01
	25	23	10.0	114	12	Q70473	Generic DNA sequence	6.58e-01
	26	23	10.0	114	12	Q70471	Generic DNA sequence	6.58e-01
c	27	23	10.0	114	12	Q70471	Generic DNA sequence	6.58e-01
	28	23	10.0	565	6	Q35072	HCV envelope region n	6.58e-01
c	29	22	9.6	36	2	Q11195	Ballast Constituent c	1.96e+00
	30	22	9.6	75	21	T13612	DC43 TSAR library gen	1.96e+00
c	31	22	9.6	81	21	T13611	DC43 TSAR library gen	1.96e+00
	32	22	9.6	82	21	T13610	DC43 TSAR library gen	1.96e+00
	33	22	9.6	89	32	T76219	Human IL5 antisense o	1.96e+00
c	34	22	9.6	91	46	V44650	Mammalian DNA replica	1.96e+00
	35	22	9.6	190	32	T76452	Chymase antisense oli	1.96e+00
	36	22	9.6	264	32	T76445	Substance P receptor	1.96e+00
c	37	22	9.6	3088	16	T05628	ADP ribosylation fact	1.96e+00
	38	21	9.2	906	51	V63027	B. malayi ankyrin cDN	5.75e+00
c	39	21	9.2	908	52	V63315	Nucleotide nBmAnk908	5.75e+00
c	40	21	9.2	911	52	V63312	Nucleotide nDiAnk911	5.75e+00
c	41	21	9.2	911	51	V63012	D. immitis ankyrin nD	5.75e+00
	42	21	9.2	1329	50	V33199	Secreted protein FB78	5.75e+00
	43	21	9.2	3871	2	N71302	HSV-1 gB and surround	5.75e+00
c	44	21	9.2	5235	51	V63022	D. immitis ankyrin nD	5.75e+00
c	45	21	9.2	5503	51	V63020	D. immitis ankyrin nD	5.75e+00

ALIGNMENTS

RESULT 1
ID V84366 standard; cDNA to mRNA; 439 BP.
AC V84366;
DT 30-MAR-1999 (first entry)
DE Human stomach carcinoma cDNA clone HP10408.
KW Transmembrane protein; HP10408; human; stomach cancer; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 75..311
FT /*tag= a
FT /note= "cDNA comprising the coding region (minus
FT the stop codon) is claimed (Claim 3)"
PN WO9855508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI; 99-045730/04.
DR P-PSDB; W88498.
PT New human proteins containing transmembrane domains and their
PT encoding sequences - useful in the preparation of antibodies and
PT large-scale protein production, gene diagnosis, and gene therapy
PS Claim 4; Page 135; 178pp; English.
CC This is the nucleotide sequence of cDNA clone HP10408, which
CC includes a coding region (also claimed) for a novel human
CC transmembrane protein (see W88498). The clone was isolated from a
CC stomach cancer cDNA library using a signal sequence detection
CC method, and by protein synthesis by in vitro translation. The
CC encoded protein has a putative signal sequence and a putative
CC internal transmembrane domain. The invention provides nucleotide
CC sequences (see V84359-76) coding for 18 transmembrane proteins

```
CC   (see W88491-508), vectors containing such polynucleotides, and
CC   eukaryotic cells containing the vectors.  The proteins can be
CC   used as antigens or as compositions in the preparation of
CC   antibodies against the proteins.  The polynucleotides can be used
CC   as probes for gene diagnosis, and as gene sources for gene therapy
CC   and large-scale production of proteins encoded by the cDNA.  The
CC   host cells are used for the detection of ligands corresponding to
CC   the expressed proteins, and the screening of low mol.wt. medicines.
SQ   Sequence   439 BP;    89 A;    137 C;    109 G;    104 T;

  Query Match           97.4%;  Score 223;  DB 60;  Length 439;
  Best Local Similarity  99.6%;  Pred. No. 8.25e-127;
  Matches   229;  Conservative    0;  Mismatches   0;  Indels   1;  Gaps   1;

Db     29 accgggacttcagtgtctcctccatcccaggagcgcagtggccactatggggtctgggct 88
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 ACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCT 60

Db     89 gccccttgtcctcctcttgaccctccttggcagctcacatggaacagggccgggtatgac 148
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 120

Db    149 tttgcaactgaagctgaaggagtcttttctgacaaattcctcctatgagtccagcttcct 208
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 TTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCT 180

Db    209 ggaattgcttgaaaagctctgcctcctcctccatctcccttcagggacca 258
          ||||||||||||||| ||||||||||||||||||||||||||||||||
Qy    181 GGAATTGCTTGAAAAG-TCTGCCTCCTCCTCCATCTCCCTTCAGGGACCA 229


RESULT    2
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match           16.2%;  Score 37;  DB 9;  Length 91;
  Best Local Similarity   5.7%;  Pred. No. 3.13e-08;
  Matches     3;  Conservative   42;  Mismatches   8;  Indels   0;  Gaps   0;

Db     11 ssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctc 63
          :::::  :: ::: :: :  :::::::::::::::::::: :::::::::|||
Cp     83 GGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTC 31


RESULT    3
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match           16.2%;  Score 37;  DB 1;  Length 204;
  Best Local Similarity   8.7%;  Pred. No. 3.13e-08;
  Matches    10;  Conservative   59;  Mismatches  46;  Indels   0;  Gaps   0;

Db     72 gaayycdchvgccgymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynach 131
          || ::|: :: |  :    |::::: :: ::::: :: |  ::   : : : :|    |:
Qy      6 GACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCC 65

Db    132 hddhyvybbbvynvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
          :::::::::::: :: :  |    :   : : :::  :: :  : :: : ::::|
Qy     66 TTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 120


RESULT    4
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
```

```
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match           16.2%;  Score 37;  DB 1;  Length 204;
  Best Local Similarity  7.9%;  Pred. No. 3.13e-08;
  Matches     8;  Conservative   54;  Mismatches  39;  Indels   0;  Gaps   0;

Db     86 ymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynv 145
          :::|   :. ::: : :: :  : : | ::| : :::::   :  |::::::::: :::  :
Cp    168 TCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGC 109

Db    146 hnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
          : :    |  :  ::|: ::    :   :::::|:::::::|
Cp    108 CCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGAC 68


RESULT    5
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match           15.7%;  Score 36;  DB 9;  Length 91;
  Best Local Similarity  0.0%;  Pred. No. 1.12e-07;
  Matches     0;  Conservative   43;  Mismatches   7;  Indels   0;  Gaps   0;

Db     11 ssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsv 60
          :::  :: ::: :::::::::::::::::::: ::::::::::::  :::
Qy      3 CGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGG 52


RESULT    6
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    5 A;    10 C;    0 G;    0 T;

  Query Match           13.5%;  Score 31;  DB 12;  Length 114;
  Best Local Similarity  6.8%;  Pred. No. 5.71e-05;
  Matches     7;  Conservative   28;  Mismatches  68;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbcacnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :     :||   :  :  :  :|||  :  :  :  :  :  :  :  :  :
Qy     18 TCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCT 77

Db     63 bnnbnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnb 105
          :  :  :  :  :  :| |     :  :  :  :  :  :  :
Qy     78 TGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 120


RESULT    7
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
```

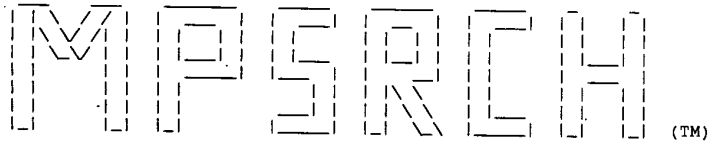
```
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           13.5%;  Score 31;  DB 12;  Length 114;
  Best Local Similarity   4.6%;  Pred. No. 5.71e-05;
  Matches     5;  Conservative   30;  Mismatches  74;  Indels   0;  Gaps   0;

Db      6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnnbnn 65
          :  :  :  :  :  :  :     :  :  :  :  :  :  :  :||   :        :
Cp    140 CCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGG 81

Db     66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnb 114
          :     :  :  :  :  :  :  :     :  :  :  :  :|||  :
Cp     80 TCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCT 32


RESULT    8
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           13.1%;  Score 30;  DB 12;  Length 114;
  Best Local Similarity   3.8%;  Pred. No. 1.93e-04;
  Matches     4;  Conservative   30;  Mismatches  72;  Indels   0;  Gaps   0;

Db      9 bnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnnbnnbnn 68
          :  :  :     :  :  :  :  : || :  :  :  :  :  :  :        :  :
Qy     15 GTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCC 74

Db     69 bnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
          :  :  :  :  :||   :  :     :  :  :  :  :  :  :
Qy     75 TCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 120


RESULT    9
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
```

```
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    5 A;    10 C;    0 G;    0 T;

  Query Match           13.1%;  Score 30;  DB 12;  Length 114;
  Best Local Similarity   6.7%;  Pred. No. 1.93e-04;
  Matches     7;  Conservative   28;  Mismatches  69;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbcacnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           :  :  :  :||   :  :  :  :| |  :  :  :     :  :  :  :
Cp     134 GCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGA 75

Db      63 bnnbnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbc 106
           :  :  :  :  :  :| |     :  :  :  :  :  :  :|
Cp      74 GGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTC 31


RESULT   10
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           13.1%;  Score 30;  DB 12;  Length 114;
  Best Local Similarity   2.7%;  Pred. No. 1.93e-04;
  Matches     3;  Conservative   32;  Mismatches  77;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           :  :  :  :     :  :  :  :  :  :  :| |  :     :  :  :  :
Cp     134 GCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGA 75

Db      63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
           :  :  :  :  :  :  :  :| |  :  :  :  :  :  :  :  :  :
Cp      74 GGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATG 23


RESULT   11
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           12.7%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   3.0%;  Pred. No. 6.43e-04;
  Matches     3;  Conservative   29;  Mismatches  69;  Indels   0;  Gaps   0;

Db       9 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnnbnnbnn 68
           :  :  :     :  :  :  :  :  :  :  :  :||   :        :  :
Qy      15 GTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCC 74
```

```
Db      69 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbt 109
           :  :  :  :  :  :  :  :  :     :  :  :  :  :|
Qy      75 TCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGT 115


RESULT   12
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           12.7%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   2.1%;  Pred. No. 6.43e-04;
  Matches     2;  Conservative   29;  Mismatches  66;  Indels   0;  Gaps   0;

Db       6 bnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnnnbnn 65
           :  :  :  :  :|| :  :  :  :  :  :  :  :  :  :  :  :  :     :
Qy      27 CCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCC 86

Db      66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 102
           :  :  :  :     :  :  :  :  :  :  :  :
Qy      87 TTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTT 123


RESULT   13
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match           12.7%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   4.6%;  Pred. No. 6.43e-04;
  Matches     5;  Conservative   29;  Mismatches  75;  Indels   0;  Gaps   0;

Db       3 cnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           |  :  :     :  :  :     :  :  : ||  :  :  :  :  :  :  :
Qy       3 CGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGC 62

Db      63 bnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 111
           :  :||   :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      63 CCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCC 111


RESULT   14
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
```

```
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           12.2%;  Score 28;  DB 12;  Length 114;
  Best Local Similarity  3.6%;  Pred. No. 2.12e-03;
  Matches     4;  Conservative   30;  Mismatches  78;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :    |  :  :  :  :  :  :  :     :  :  :
Cp    137 TCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCA 78

Db     63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnb 114
            :  :  :  :  :  :  :  :  :  :  :  :  :|||  :  :  :
Cp     77 AGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGG 26


RESULT   15
ID   T76405 standard; DNA; 178 BP.
AC   T76405;
DT   15-SEP-1997  (first entry)
DE   Human endothelin-1 antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Claim 5; Page 38; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human endothelin-1, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   178 BP;    0 A;    52 C;    46 G;    32 T;

  Query Match           12.2%;  Score 28;  DB 32;  Length 178;
  Best Local Similarity  25.0%;  Pred. No. 2.12e-03;
  Matches    18;  Conservative   32;  Mismatches  22;  Indels   0;  Gaps   0;

Db    100 cttggcbbgccbcbbbcbgcbgbgbgbbbbtcbtgbgcbbbtbbtccbttctgbbbbbbb 159
          || |||:: ||: :::|: |: : :| :::||:| :|| ::|:    :     :::::::
Qy     54 CTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGG 113

Db    160 gggbtcbbbbbc 171
          |  :  :::::|
Qy    114 GTATGACTTTGC 125


Search completed: Sun Oct 24 17:08:37 1999
Job time : 76 secs.
```

****



****

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 17:08:54 1999; MasPar time 20.47 Seconds
967.578 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-2
Description: (1-229) from US09092296.seq
Perfect Score: 229
N.A. Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCCTTCAGGGACCA 229
Comp: TGGCCCTGAAGTCACAGAGG.....GGTAGAGGGAAGTCCCTGGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-issued
1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 7.362; Variance 4.165; scale 1.768

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	46	20.1	7218	2	US-08-232-	Sequence 14, Applicati	3.37e-16
c	2	33	14.4	215	1	US-08-238-	Sequence 5, Applicatio	6.36e-08
	3	31	13.5	965	3	US-08-388-	Sequence 22, Applicati	1.05e-06
	4	27	11.8	215	1	US-08-238-	Sequence 5, Applicatio	2.44e-04
c	5	27	11.8	965	3	US-08-388-	Sequence 22, Applicati	2.44e-04
	6	22	9.6	75	4	PCT-US95-1	Sequence 99, Applicati	1.54e-01
c	7	22	9.6	81	4	PCT-US95-1	Sequence 98, Applicati	1.54e-01
	8	22	9.6	82	4	PCT-US95-1	Sequence 97, Applicati	1.54e-01
c	9	22	9.6	3088	3	US-08-418-	Sequence 1, Applicatio	1.54e-01
c	10	21	9.2	65	1	US-08-471-	Sequence 145, Applicat	5.27e-01
	11	21	9.2	66	1	US-08-471-	Sequence 144, Applicat	5.27e-01
c	12	21	9.2	68	1	US-08-471-	Sequence 143, Applicat	5.27e-01
	13	21	9.2	69	1	US-08-471-	Sequence 142, Applicat	5.27e-01
	14	21	9.2	74	4	PCT-US95-1	Sequence 94, Applicati	5.27e-01
c	15	21	9.2	74	4	PCT-US95-1	Sequence 100, Applicat	5.27e-01
c	16	21	9.2	75	4	PCT-US95-1	Sequence 99, Applicati	5.27e-01
	17	21	9.2	81	4	PCT-US95-1	Sequence 98, Applicati	5.27e-01
	18	21	9.2	81	4	PCT-US95-1	Sequence 92, Applicati	5.27e-01
c	19	21	9.2	82	4	PCT-US95-1	Sequence 97, Applicati	5.27e-01
c	20	21	9.2	92	3	US-08-353-	Sequence 16, Applicati	5.27e-01
	21	21	9.2	906	3	US-09-031-	Sequence 41, Applicati	5.27e-01
	22	21	9.2	906	3	US-08-847-	Sequence 41, Applicati	5.27e-01
c	23	21	9.2	906	3	US-08-847-	Sequence 40, Applicati	5.27e-01
c	24	21	9.2	906	3	US-09-031-	Sequence 40, Applicati	5.27e-01
	25	21	9.2	908	3	US-09-031-	Sequence 39, Applicati	5.27e-01
	26	21	9.2	908	3	US-08-847-	Sequence 39, Applicati	5.27e-01
c	27	21	9.2	908	3	US-08-847-	Sequence 37, Applicati	5.27e-01
c	28	21	9.2	908	3	US-09-031-	Sequence 37, Applicati	5.27e-01
	29	21	9.2	909	3	US-09-031-	Sequence 26, Applicati	5.27e-01
	30	21	9.2	909	3	US-08-847-	Sequence 26, Applicati	5.27e-01
c	31	21	9.2	909	3	US-09-031-	Sequence 25, Applicati	5.27e-01
c	32	21	9.2	909	3	US-08-847-	Sequence 25, Applicati	5.27e-01
	33	21	9.2	911	3	US-09-031-	Sequence 24, Applicati	5.27e-01
	34	21	9.2	911	3	US-08-847-	Sequence 24, Applicati	5.27e-01
c	35	21	9.2	911	3	US-08-847-	Sequence 22, Applicati	5.27e-01
c	36	21	9.2	911	3	US-09-031-	Sequence 22, Applicati	5.27e-01
	37	21	9.2	5235	3	US-09-031-	Sequence 36, Applicati	5.27e-01
	38	21	9.2	5235	3	US-08-847-	Sequence 36, Applicati	5.27e-01
c	39	21	9.2	5235	3	US-08-847-	Sequence 35, Applicati	5.27e-01
c	40	21	9.2	5235	3	US-09-031-	Sequence 35, Applicati	5.27e-01
	41	21	9.2	5503	3	US-09-031-	Sequence 34, Applicati	5.27e-01
	42	21	9.2	5503	3	US-08-847-	Sequence 34, Applicati	5.27e-01
c	43	21	9.2	5503	3	US-08-847-	Sequence 32, Applicati	5.27e-01
c	44	21	9.2	5503	3	US-09-031-	Sequence 32, Applicati	5.27e-01
c	45	20	8.7	66	4	PCT-US95-1	Sequence 93, Applicati	1.76e+00

ALIGNMENTS

RESULT 1
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC xxxxxx
DT
DE Sequence 14, Application US/08232463
CC Sequence 14, Application US/08232463
CC Patent No. 5670367
CC GENERAL INFORMATION:
CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/232,463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/935,313
CC FILING DATE:
CC APPLICATION NUMBER: EP 91 114 300.6
CC FILING DATE: 26-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.
CC REGISTRATION NUMBER: 29,768
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)683-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:

```
CC        LENGTH:  7218 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      IMMEDIATE SOURCE:
CC        CLONE:  pTZgpt-F1s
SQ   SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

  Query Match           20.1%;  Score 46;  DB 2;  Length 7218;
  Best Local Similarity   2.3%;  Pred. No. 3.37e-16;
  Matches     5;  Conservative  129;  Mismatches  88;  Indels   0;  Gaps   0;

Db   1219 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1278
           ::::  : :::::::: ::::     : :  :  :: :: :    :::   :: ::::::
Qy      8 CTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTT 67

Db   1279 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1338
           ::::::::::  ::::::::  :  ::: : :    :    ::   : :  :::: :
Qy     68 GTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAA 127

Db   1339 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1398
           ::   ::      ::::::::  :  :::::::: :  :::  ::::::    ::
Qy    128 CTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTG 187

Db   1399 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440
           :::      ::: :::::::::: :::::::::  | ||||
Qy    188 CTTGAAAAGTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCA 229


RESULT    2
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC     INFORMATION FOR SEQ ID NO:  5:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  215 base pairs
CC         TYPE:  nucleic acid
CC         STRANDEDNESS:  single
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  protein
CC       FEATURE:
CC         NAME/KEY:  misc_feature
CC         LOCATION:  1..215
CC         OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC         OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           14.4%;  Score 33;  DB 1;  Length 215;
  Best Local Similarity  14.6%;  Pred. No. 6.36e-08;
  Matches    26;  Conservative   73;  Mismatches  77;  Indels   2;  Gaps   2;

Db       1 MTNVTMSSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNK 60
             :| : :::  :: : |:  ::|:::|     : |    : :|:|:  ||:::  :: : :
Cp     185 ATTCCAGGAAGCTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTC-AGT 127

Db      61 YSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTD-GNRSGADS 119
             :::|   :    :   : ::: :| :   ::  :|::: |:| :::    : | ::||
Cp     126 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 67

Db     120 YGSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTKKHAKNSADGKVGSKNNGDR 177
           |::: |  :  :       |:::: :   ::|::  |:: |: :|:| : :   |::
Cp      66 AGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATGGAGGAGACACTGAA 9


RESULT    3
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
CC        TELEFAX:  212-838-3884
CC     INFORMATION FOR SEQ ID NO:  22:
CC       SEQUENCE CHARACTERISTICS:
CC         LENGTH:  965 base pairs
CC         TYPE:  nucleic acid
CC         STRANDEDNESS:  unknown
CC         TOPOLOGY:  unknown
CC       MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.
```

```
  Query Match           13.5%;  Score 31;  DB 3;  Length 965;
  Best Local Similarity  19.3%;  Pred. No. 1.05e-06;
  Matches    23;  Conservative   53;  Mismatches  42;  Indels   1;  Gaps   1;

Db    845 SVTAADTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHUVKDMTSSSSASVGDRVTTCRSST 904
          ::|  :| ::  ::| ::::::|| ::   |: :: ::::::|::::   |: :| | : |
Qy     19 CCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTT 78

Db    905 THGNGNTYYWYKGKAKYRVSNRSGVSRSGSGSGTDYTTSSDATYYCGTHARTGGTKVKG 963
            :    |:::::| |:: :  : |: :: :|:  :  |  :| :: | :| ||  :::|
Qy     79 GACC-CTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG 136


RESULT    4
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC    INFORMATION FOR SEQ ID NO:  5:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  215 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  protein
CC      FEATURE:
CC        NAME/KEY:  misc_feature
CC        LOCATION:  1..215
CC        OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC        OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           11.8%;  Score 27;  DB 1;  Length 215;
  Best Local Similarity  20.0%;  Pred. No. 2.44e-04;
  Matches    22;  Conservative   42;  Mismatches  45;  Indels   1;  Gaps   1;

Db     18 CNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAK 77
          | :: : :| |  ::| :   :  |:|: :| :::  : |:  : :|   :|   :|| :
Qy     85 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG-AAGGAGT 143

Db     78 THYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
          :::|:| :: |: : :| ::  :   :: | |:   :||: :|: : |:
Qy    144 CTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAA 193


RESULT    5
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
CC        TELEFAX:  212-838-3884
CC    INFORMATION FOR SEQ ID NO:  22:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  965 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  unknown
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

  Query Match           11.8%;  Score 27;  DB 3;  Length 965;
  Best Local Similarity  15.6%;  Pred. No. 2.44e-04;
  Matches    22;  Conservative   59;  Mismatches  60;  Indels   0;  Gaps   0;

Db    790 GGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKRGRVTMADTSSNSRSSVTAA 849
          ||::: :| | ::     :  ::|: :| :|| :|    :: :: : :|:   :  :  |
Cp    202 GGCAGACTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGGAGGAATTTGTCAGAAAAGA 143

Db    850 DTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHUVKDMTSSSSASVGDRVTTCRSSTTHGNG 909
           | :::|::   :::: |  : || :|:::: :  :|:  :   |:::   :::   | |
Cp    142 CTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAG 83

Db    910 NTYYWYKGKAKYRVSNRSGVS 930
            ::: : :   ::: ::|::
Cp     82 GGTCAAGAGGAGGACAAGGGG 62


RESULT    6
```

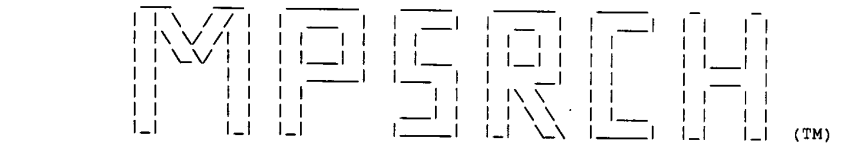
```
ID   PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.
AC   xxxxxx
DT
DE   Sequence 99, Application PC/TUS9511934
CC   Sequence 99, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  99:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  75 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 75 BP; 1 A; 1 C; 7 G; 5 T; 61 OTHER.

  Query Match             9.6%;  Score 22;  DB 4;  Length 75;
  Best Local Similarity   8.7%;  Pred. No. 1.54e-01;
  Matches     6;  Conservative   19;  Mismatches  44;  Indels   0;  Gaps   0;

Db       5 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 64
           |   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :     :  :  :
Qy      50 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 109

Db      65 NBGGTTGTG 73
            :|| | ||
Qy     110 CCGGGTATG 118


RESULT    7
ID   PCT-US95-11934-98 STANDARD; DNA; UNC; 81 BP.
AC   xxxxxx
DT
DE   Sequence 98, Application PC/TUS9511934
CC   Sequence 98, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  98:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  81 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 81 BP; 6 A; 6 C; 4 G; 5 T; 60 OTHER.

  Query Match             9.6%;  Score 22;  DB 4;  Length 81;
  Best Local Similarity   7.6%;  Pred. No. 1.54e-01;
  Matches     5;  Conservative   20;  Mismatches  41;  Indels   0;  Gaps   0;

Db       5 TGTCTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNV 64
           || | ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Cp      92 TGCCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGC 33

Db      65 NNVNNV 70
             :  :
Cp      32 TCCTGG 27


RESULT    8
ID   PCT-US95-11934-97 STANDARD; DNA; UNC; 82 BP.
AC   xxxxxx
DT
DE   Sequence 97, Application PC/TUS9511934
CC   Sequence 97, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
```

```
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  97:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  82 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 82 BP; 1 A; 2 C; 10 G; 8 T; 61 OTHER.

  Query Match             9.6%;  Score 22;  DB 4;  Length 82;
  Best Local Similarity   8.7%;  Pred. No. 1.54e-01;
  Matches     6;  Conservative   19;  Mismatches  44;  Indels   0;  Gaps   0;

Db      12 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 71
           |   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :     :  :  :
Qy      50 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 109

Db      72 NBGGTTGTG 80
             :|| | ||
Qy     110 CCGGGTATG 118


RESULT    9
ID   US-08-418-444A-1 STANDARD; DNA; UNC; 3088 BP.
AC   xxxxxx
DT
DE   Sequence 1, Application US/08418444A
CC   Sequence 1, Application US/08418444A
CC   Patent No. 5773688
CC    GENERAL INFORMATION:
CC      APPLICANT:  KURODA, HISAO
CC      APPLICANT:  HIROTA, NAOHIKO
CC      APPLICANT:  ITO, KAZUTOSHI
CC      TITLE OF INVENTION:  GENE EXPRESSION VECTOR USING THE GENE
CC      TITLE OF INVENTION:  EXPRESSION REGULATING REGION OF THE ADP RIBOSYLATION
CC      TITLE OF INVENTION:  FACTOR
CC      NUMBER OF SEQUENCES:  9
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
CC        STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC        CITY:  ARLINGTON
CC        STATE:  VIRGINIA
CC        COUNTRY:  USA
CC        ZIP:  22202
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/418,444A
CC        FILING DATE:  07-APR-1995
CC        CLASSIFICATION:  800
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  JP HEI 6-71048
CC        FILING DATE:  08-APR-1994
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Oblon, No. 5773688man F.
CC        REGISTRATION NUMBER:  24,618
CC        REFERENCE/DOCKET NUMBER:  2589-024-0
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703) 413-3000
CC        TELEFAX:  (703) 413-2220
CC        TELEX:  248855 OPAT UR
CC    INFORMATION FOR SEQ ID NO:  1:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  3088 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  double
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 3088 BP; 716 A; 761 C; 672 G; 939 T; 0 OTHER.

  Query Match             9.6%;  Score 22;  DB 3;  Length 3088;
  Best Local Similarity  72.0%;  Pred. No. 1.54e-01;
  Matches    36;  Conservative    0;  Mismatches  14;  Indels   0;  Gaps   0;

Db      15 CCGAAGATGGGCCAGAGGGGGGCCAGGGGCTGCCCAGGCGCCCTGCTGGC 64
           || | || || | ||||| || | |||||| |||||| | || |  ||||
Cp      90 CCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGC 41


RESULT   10
ID   US-08-471-052A-145 STANDARD; DNA; UNC; 65 BP.
AC   xxxxxx
DT
DE   Sequence 145, Application US/08471052A
CC   Sequence 145, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  145:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  65 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 65 BP; 3 A; 3 C; 3 G; 2 T; 54 OTHER.

  Query Match             9.2%;  Score 21;  DB 1;  Length 65;
  Best Local Similarity  13.8%;  Pred. No. 5.27e-01;
  Matches     9;  Conservative   16;  Mismatches  40;  Indels   0;  Gaps   0;

Db       1 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNA 60
           || ||:  :  :  :  :  :     :  :  :  :  :  :  :  :  :  :  :  |
Cp     174 CTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATA 115

Db      61 CCTGG 65
           || ||
Cp     114 CCCGG 110


RESULT   11
ID   US-08-471-052A-144 STANDARD; DNA; UNC; 66 BP.
AC   xxxxxx
DT
```

```
DE   Sequence 144, Application US/08471052A
CC   Sequence 144, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  144:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  66 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 66 BP; 2 A; 3 C; 4 G; 2 T; 55 OTHER.

  Query Match             9.2%;  Score 21;  DB 1;  Length 66;
  Best Local Similarity   5.2%;  Pred. No. 5.27e-01;
  Matches     3;  Conservative   18;  Mismatches  37;  Indels   0;  Gaps   0;

Db       4 AGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNC 61
           ||   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy      37 AGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGC 94


RESULT   12
ID   US-08-471-052A-143 STANDARD; DNA; UNC; 68 BP.
AC   xxxxxx
DT
DE   Sequence 143, Application US/08471052A
CC   Sequence 143, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  143:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  68 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 68 BP; 3 A; 3 C; 5 G; 3 T; 54 OTHER.

  Query Match             9.2%;  Score 21;  DB 1;  Length 68;
  Best Local Similarity  13.8%;  Pred. No. 5.27e-01;
  Matches     9;  Conservative   16;  Mismatches  40;  Indels   0;  Gaps   0;

Db       4 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNA 63
           || ||:  :  :  :  :     :  :  :  :  :  :  :  :     :  :  :  |
Cp     174 CTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATA 115

Db      64 CCTGG 68
           || ||
Cp     114 CCCGG 110


RESULT   13
ID   US-08-471-052A-142 STANDARD; DNA; UNC; 69 BP.
AC   xxxxxx
DT
DE   Sequence 142, Application US/08471052A
CC   Sequence 142, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  142:
CC      SEQUENCE CHARACTERISTICS:
```

```
CC        LENGTH:  69 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 69 BP; 2 A; 4 C; 6 G; 2 T; 55 OTHER.

  Query Match             9.2%;  Score 21;  DB 1;  Length 69;
  Best Local Similarity   5.2%;  Pred. No. 5.27e-01;
  Matches     3;  Conservative   18;  Mismatches  37;  Indels   0;  Gaps   0;

Db       7 AGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBC 64
           ||   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy      37 AGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGC 94


RESULT   14
ID   PCT-US95-11934-94 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 94, Application PC/TUS9511934
CC   Sequence 94, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  94:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 3 A; 4 C; 3 G; 1 T; 63 OTHER.

  Query Match             9.2%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity   1.7%;  Pred. No. 5.27e-01;
  Matches     1;  Conservative   20;  Mismatches  39;  Indels   0;  Gaps   0;

Db       9 BNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNA 68
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy      33 GCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCA 92


RESULT   15
ID   PCT-US95-11934-100 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 100, Application PC/TUS9511934
CC   Sequence 100, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  100:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 6 A; 6 C; 1 G; 1 T; 60 OTHER.

  Query Match             9.2%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity   7.5%;  Pred. No. 5.27e-01;
  Matches     5;  Conservative   19;  Mismatches  43;  Indels   0;  Gaps   0;

Db       5 AVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNN 64
           |:  :  :  :  :  :  :  :     :  :  :  :  :  :  :  :  :  :  :  :
Cp     121 AGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 62

Db      65 NACCACA 71
           | | ||
Cp      61 CAGCCCA 55


Search completed: Sun Oct 24 17:09:17 1999
Job time : 23 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 16:59:27 1999; MasPar time 446.22 Seconds
1202.498 Million cell updates/sec
Tabular output not generated.

Title: >US-09-092-296-2
Description: (1-229) from US09092296.seq
Perfect Score: 229
N.A. Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCCTTCAGGGACCA 229
Comp: TGGCCCTGAAGTCACAGAGG.....GGTAGAGGGAAGTCCCTGGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est2 20:gb_est20
21:gb_est21 22:gb_est22 23:gb_est23 24:gb_est24
25:gb_est25 26:gb_est26 27:gb_est27 28:gb_est28
29:gb_est29 30:gb_est3 31:gb_est4 32:gb_est5 33:gb_est6
34:gb_est7 35:gb_est8 36:gb_est9 37:gb_gss1 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 9.755; Variance 2.089; scale 4.670

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
c	1	53	23.1	328	23	AI136523	UI-R-C2p-nq-e-02-0-UI.	4.59e-50
	2	52	22.7	252	17	AA754459	97SN1787 Rice Immature	2.11e-48
c	3	50	21.8	252	17	AA754459	97SN1787 Rice Immature	4.21e-45
c	4	40	17.5	247	17	AA754458	97SN1784 Rice Immature	4.11e-29
	5	32	14.0	247	17	AA754458	97SN1784 Rice Immature	3.60e-17
c	6	32	14.0	2275	20	AF034173	AF034173 Human mRNA (T	3.60e-17
	7	27	11.8	2275	20	AF034173	AF034173 Human mRNA (T	2.78e-10
	8	24	10.5	348	19	F06958	HSC1QC101 normalized i	1.90e-06
	9	23	10.0	238	12	AA376266	EST88915 HSC172 cells	3.12e-05
	10	23	10.0	311	11	AA323964	EST26816 Cerebellum II	3.12e-05
	11	23	10.0	364	34	W33870	mc56b03.r1 Soares mous	3.12e-05
c	12	23	10.0	412	18	AA769782	ah71b05.s1 Soares_test	3.12e-05
	13	23	10.0	426	8	T15255	crs855 lambdaZAPST Ric	3.12e-05
	14	23	10.0	481	14	AA498488	vh44b06.r1 Barstead mo	3.12e-05
	15	23	10.0	1287	20	AF038250	AF038250 Human mRNA (T	3.12e-05
	16	22	9.6	308	8	D40392	RICS2342A Rice shoot O	4.75e-04
	17	22	9.6	339	19	F08745	HSC1DB011 normalized i	4.75e-04
c	18	22	9.6	395	41	AQ375965	RPCI11-161H11.TV RPCI1	4.75e-04
c	19	22	9.6	404	17	AA730403	nw42c07.s1 NCI_CGAP_Ew	4.75e-04
	20	22	9.6	433	34	W79098	zd75h10.r1 Soares_feta	4.75e-04
	21	22	9.6	441	14	C28493	C28493 Rice callus cDN	4.75e-04
c	22	22	9.6	634	22	AI063013	GH02423.5prime GH Dros	4.75e-04
	23	22	9.6	1025	37	B12587	F22O11-Sp6.1 IGF Arabi	4.75e-04
c	24	21	9.2	299	8	M79528	wEST00065 Mixed stage,	6.60e-03
	25	21	9.2	370	25	AI328672	a5g11ne.r1 Neurospora	6.60e-03
	26	21	9.2	384	23	AI160316	qb66d04.x1 Soares_feta	6.60e-03
c	27	21	9.2	401	42	AQ444216	GSSTc0313 Trypanosoma	6.60e-03
	28	21	9.2	417	30	R36249	yh91f09.r1 Soares plac	6.60e-03
	29	21	9.2	420	28	AI566143	tn53h10.x1 NCI_CGAP_Ki	6.60e-03
c	30	21	9.2	422	10	AA275959	vc27e03.r1 Barstead MP	6.60e-03
c	31	21	9.2	429	28	AI509530	vx09h04.y1 Soares 2NbM	6.60e-03
c	32	21	9.2	440	18	AA822673	vw14f01.r1 Soares 2NbM	6.60e-03
c	33	21	9.2	471	40	AQ236981	RPCI11-64I22.TK RPCI11	6.60e-03
	34	21	9.2	494	20	AA898724	NCM6E8T7 Mycelial Neur	6.60e-03
	35	21	9.2	499	37	B45020	HS-1060-B1-B01-MF.abi	6.60e-03
c	36	21	9.2	502	15	AA586074	28723 Lambda-PRL2 Arab	6.60e-03
c	37	21	9.2	524	10	AA239967	mw24g07.r1 Soares mous	6.60e-03
c	38	21	9.2	533	15	AA593745	nl83g11.s1 NCI_CGAP_Br	6.60e-03
	39	21	9.2	550	20	AA898701	NCM6G7T7 Mycelial Neur	6.60e-03
c	40	21	9.2	567	30	R61539	yh16f01.s1 Soares infa	6.60e-03
	41	21	9.2	594	20	AA898768	NCP4C7T7 Perithecial N	6.60e-03
c	42	21	9.2	601	32	N37023	yy40d03.s1 Soares mela	6.60e-03
c	43	21	9.2	688	37	AG015068	Homo sapiens genomic D	6.60e-03
c	44	21	9.2	694	36	AA140679	CK00462.5prime CK Dros	6.60e-03
c	45	21	9.2	796	37	AF035986	Salmonella typhimurium	6.60e-03

ALIGNMENTS

RESULT 1
LOCUS AI136523 328 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C2p-nq-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone UI-R-C2p-nq-e-02-0-UI 3', mRNA sequence.
ACCESSION AI136523
NID g3637300
VERSION AI136523.1 GI:3637300
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 328)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 14, 1998 this sequence version replaced gi:1877567.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.
FEATURES Location/Qualifiers

```
    source          1. .328
                    /organism="Rattus norvegicus"
                    /strain="Sprague-Dawley"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p
                    library is a subtracted library derived from the UI-R-C1
                    library, which is a subtracted library derived from the
                    UI-R-C0 library.  The UI-R-C0 library consisted of a
                    mixture of individually tagged normalized  libraries
                    constructed from rat placenta, adult lung, brain, liver,
                    kidney,  heart, spleen, ovary, muscle, 8, 12 and 18-day
                    embryo. The tag is a string of  3-5 nucleotides present
                    between the Not I site and the oligo-dT track which allows
                    identification of the library of origin of a clone within
                    the mixture. The subtracted library (UI-R-C2p) was
                    constructed as follows: PCR amplified cDNA inserts from
                    UI-R-C1 clones from which 3' ESTs had been derived was
                    used as a driver in a hybridization with the UI-R-C1
                    library in the form of single-stranded circles. The
                    remaining single-stranded circles (subtracted library) was
                    purified by hydroxyapatite column chromatography,
                    converted to double-stranded circles and electroporated
                    into DH10B bacteria (Life Technologies) to generate the
                    UI-R-C2p library. This procedure has been previously
                    described (Bonaldo, Lennon and Soares, Genome Research 6:
                    791-806, 1996)"
                    /db_xref="taxon:10116"
                    /clone="UI-R-C2p-nq-e-02-0-UI"
                    /clone_lib="UI-R-C2p"
                    /dev_stage="adult"
                    /lab_host="DH10B (Life Technologies)"
BASE COUNT       62 a     77 c     98 g     91 t
ORIGIN

  Query Match           23.1%;  Score 53;  DB 23;  Length 328;
  Best Local Similarity  75.4%;  Pred. No. 4.59e-50;
  Matches    86;  Conservative    0;  Mismatches  27;  Indels   1;  Gaps   1;

Db    187 TGGTCCCCGGTGAGAGGTGGAGGAGGAGGCAGATCTTTTGGAGCATGTCCAGAAAGCCGG 246
        ||||||| |  | ||| ||||||||||||||||| |||||  ||||  ||||| |||| ||
Cp    229 TGGTCCCTGAAGGGAGATGGAGGAGGAGGCAGA-CTTTTCAAGCAATTCCAGGAAGCTGG 171

Db    247 AGTCTTGGGAGGTCTTGGCCTGAGATGCCTCGGTCAGTTTTACTTGCAAAGTCA 300
        | || | |||||  || | | || | | |||  |||| || | ||||||||||
Cp    170 ACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCA 117


RESULT    2
LOCUS       AA754459      252 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1787, mRNA sequence.
ACCESSION   AA754459
NID         g2801165
VERSION     AA754459.1  GI:2801165
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797457.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES            Location/Qualifiers
     source         1. .252
                    /organism="Oryza sativa"
                    /cultivar="Milyang23"
                    /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                    XhoI; Directional cDNA library inserted into lambda ZAPII
                    vector  at 5'end with EcoRI and 3' end with Xho I site."
                    /db_xref="taxon:4530"
                    /map="6"
                    /clone="97SN1787"
                    /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                    /tissue_type="Immature Seed"
                    /dev_stage="5 days after pollination"
                    /lab_host="E. coli SOLR"
BASE COUNT       5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match           22.7%;  Score 52;  DB 17;  Length 252;
  Best Local Similarity  11.9%;  Pred. No. 2.11e-48;
  Matches    20;  Conservative   87;  Mismatches  59;  Indels   2;  Gaps   2;

Db     23 YBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWBV 82
          ::|:  :: : :|::| :: ::   ::||: : :    : :|:||: ::  : : :: :
Qy     25 TCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCT 84

Db     83 BNTKVDVGNHTRCSRWRBVTRMAH-YHDYTNCBBYNNNDYHMWHBBMYBBTGCMTCTMWC 141
          : |:::: :| :::::: :  :::    ::    : ::::::: ::|| :    :|
Qy     85 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTC 144

Db    142 WBHYNTKCTASGWHTSTNYDVKS-STNTWGVTBSYDKSMHGYWCSBBV 188
          :::: |:  |  :: :| ::::: :|  :|:|:::::::: : | :::
Qy    145 TTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGA 192


RESULT    3
LOCUS       AA754459      252 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1787, mRNA sequence.
ACCESSION   AA754459
NID         g2801165
VERSION     AA754459.1  GI:2801165
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797457.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES            Location/Qualifiers
     source         1. .252
                    /organism="Oryza sativa"
```

```
                     /cultivar="Milyang23"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
                     /db_xref="taxon:4530"
                     /map="6"
                     /clone="97SN1787"
                     /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
BASE COUNT        5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match           21.8%;  Score 50;  DB 17;  Length 252;
  Best Local Similarity   9.8%;  Pred. No. 4.21e-45;
  Matches    21;  Conservative  108;  Mismatches  83;  Indels   2;  Gaps   2;

Db      22 SYBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWB 81
           ::: :  ::::   ::|   ::  | :: |: :|  : : |     : ::  :|: ::::
Cp     224 CCTGAAGGGAGATGGAGGAGGAGGCAGACTTTTCAAGCAATTCCAGGAAGCTGGACTCAT 165

Db      82 VBNTKVDVGNHTRCSRWRBVTRMAHYHDYTNCBBYNNNDYHMWHBBMYBBTGCMTCTMWC 141
           ::  ::::  :  | ::: : :: ::::: |:::   : ::::::: :   |  | ::
Cp     164 AGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTG 105

Db     142 WBHYNTKCTASGWHTSTNYDVKSSTNTWGVTBSYDKSMHGYW-C-SBBVKYHTKVSTTRA 199
           :::: |:  |: :  :   :::::|  :|: :  : :::|   | ::::  :  :: |:
Cp     104 TTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAG 45

Db     200 TRSYTCVRKYCVMWMTKKVVKKYHVVBBGCHBTD 233
           |:::  : :: :  :|:::::: ::: : |::|:
Cp      44 TGGCCACTGCGCTCCTGGGATGGAGGAGACACTG 11


RESULT    4
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
VERSION     AA754458.1  GI:2801164
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797455.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
                     /db_xref="taxon:4530"
                     /map="6"
                     /clone="97SN1784"
                     /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match           17.5%;  Score 40;  DB 17;  Length 247;
  Best Local Similarity  10.7%;  Pred. No. 4.11e-29;
  Matches    22;  Conservative  100;  Mismatches  81;  Indels   3;  Gaps   3;

Db      34 VRRVGTTTNNGKHNGRTTTWNDCSDNAHCRYTVBWYYARSKYGYGTBYYSWNVDTNTGGT 93
           ::::| |   |:: |:   : :| :  : :  : ::: :::    |:  :: ::|  |
Cp     218 GGGAGATGGAGGAGGAGGCAGACTTTTCAAGCAATTCCAGGAAGCTGGACT-CATAGGAG 160

Db      94 GVGKTTVNVHSGWNNRCSNSVVYVWBTAYCDYBHYBDRANHVDDTRCTNDRGYCNYTASD 153
           |: :||: ::: :  : : :: ::::  :|: :: :::| ::::  |  :  :  :| ::
Cp     159 GAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCA 100

Db     154 NGTSAT-KRVTGYDKTDSDCGGGCWRKVTYGSS-BYBRCGVNVMVRTTSMWTDKSTKMBS 211
            ||:|  :::   :: :::     ::::   :  : ::| : ::::  ::  : :|: ::
Cp      99 TGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCC 40

Db     212 MDMSRRSRVHYGRWMBNKKRGMSRNW 237
           :  : :: :::|: :: : :|::: :
Cp      39 ACTGCGCTCCTGGGATGGAGGAGACA 14


RESULT    5
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
VERSION     AA754458.1  GI:2801164
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797455.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
                     /db_xref="taxon:4530"
                     /map="6"
                     /clone="97SN1784"
                     /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
```

```
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
BASE COUNT         7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match           14.0%;  Score 32;  DB 17;  Length 247;
  Best Local Similarity  13.3%;  Pred. No. 3.60e-17;
  Matches    26;  Conservative   88;  Mismatches  79;  Indels   3;  Gaps   3;

Db     48 GRTTTWNDCSDNAHCRYTVBWYYARS-KYGYGTBYYSWNVDTNTGGTGVGKTTVNVHSGW 106
          |: ||  : ::  :| : :: ::|:: ::|:  :  :  ::   ||  : :  : :::
Qy      6 GACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCC 65

Db    107 NNRCSNSVVYVWBTAYCDYBHYBDRANHVDDTRCTNDRGYCNYTASDNGTSATKRVTGYD 166
          : :  :::::: |:| :::::::   ::    | :: :     :: || ||:::| ::
Qy     66 TTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGT-ATGACTTTG 124

Db    167 KTDSDCGGGCWRKVTYGS-SBYBRCGVNVMVRTTSMWTDKSTKMBSMDMSRRSRVHYGRW 225
          :::   ||:: :   : :::: | : ::::||:::  : |::: : ::    :: |::
Qy    125 CAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAA 184

Db    226 MBNKKRGMSRNWTDTK 241
          :  : |: :  | |:
Qy    185 TTGCTTGAAAAGTCTG 200


RESULT    6
LOCUS       AF034173     2275 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
VERSION     AF034173.1  GI:2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045115.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT       438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match           14.0%;  Score 32;  DB 20;  Length 2275;
  Best Local Similarity   9.6%;  Pred. No. 3.60e-17;
  Matches    11;  Conservative   65;  Mismatches  38;  Indels   1;  Gaps   1;

Db   1480 RYKRWKRRKGRRKRMTGMYKRMYRAMMMAMCAMMACWWYYWKMRGMKKCWKYRKYKKYTS 1539
          ::::::::::|::: :  :::  ::| ::|   ::  :: ::: :     :: ::   :|
Cp    213 ATGGAGGAGGAGGCAGACTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGGAGGAATTT 154

Db   1540 TYYKSWSRWYWYTTYTYWYCWCCTSMKSASCAMMRWMGYMGSRSSRSYWGYWGSM 1594
          :: :: :: ::| :|: :  | |:::  :||::::: :: : :: :::|:: ::
Cp    153 GTCAGAAAAGACTCCTT-CAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCA 100


RESULT    7
LOCUS       AF034173     2275 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
VERSION     AF034173.1  GI:2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045115.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT       438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match           11.8%;  Score 27;  DB 20;  Length 2275;
  Best Local Similarity  16.2%;  Pred. No. 2.78e-10;
  Matches    11;  Conservative   39;  Mismatches  17;  Indels   1;  Gaps   1;

Db   1556 YWYCWCCTSMKSASCAMMRWMGYMGS-RSSRSYWGYWGSMSGCYGMTKRYYRYSWTGWTK 1614
          :::|: | ::::|:| :::: |:: : : ::: : : |:: ||: :|:::: :: | :|:
Qy     20 CTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTG 79

Db   1615 WTTWWYMW 1622
          :  ::::
Qy     80 ACCCTCCT 87


RESULT    8
LOCUS       F06958       348 bp    mRNA            EST       20-FEB-1995
DEFINITION  HSC1QC101 normalized infant brain cDNA Homo sapiens cDNA clone
            c-1qc10, mRNA sequence.
ACCESSION   F06958
NID         g672595
VERSION     F06958.1  GI:672595
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 348)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     On Sep 21, 1992 this sequence version replaced gi:279286.
```

```
            Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-1qc10
            Insert Length: 639   Std Error: 0.00
            Seq primer: (-21)M13_universal
            High quality sequence stop: 150.
FEATURES              Location/Qualifiers
     source           1. .348
                      /organism="Homo sapiens"
                      /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                      Site_2: NotI; sex=Female; dev_stage=3 months old;
                      isolate=muscular atrophy patient;  tissue_type=total
                      brain;  total mRNA was oligo-(dT) primed and directionally
                      cloned 5' -> 3' into the  HindIII -> NotI sites of the
                      lafmid BA vector. Clone library from B.Soares,  Psychiatry
                      Dept. Columbia University, USA.  Normalization_method:
                      Bento Soares, P.N.A.S in press"
                      /db_xref="taxon:9606"
                      /clone="c-1qc10"
                      /clone_lib="normalized infant brain cDNA"
                      /sex="Female"
                      /tissue_type="total brain"
                      /dev_stage="3 months old"
BASE COUNT       84 a     91 c     90 g     81 t      2 others
ORIGIN

  Query Match           10.5%;  Score 24;  DB 19;  Length 348;
  Best Local Similarity  67.2%;  Pred. No. 1.90e-06;
  Matches    45;  Conservative    0;  Mismatches  22;  Indels   0;  Gaps   0;

Db       5 TGTTGCCTCCCTCCCAGGATCCCTTTGGTGAGTATGGTGTTCAGGNTGCACCACCACCAC 64
           |||  ||||| |||||||| | |  |||  | ||||| ||   || ||| ||    || |
Qy      14 TGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTC 73

Db      65 CTCTAGA 71
           |||| ||
Qy      74 CTCTTGA 80


RESULT    9
LOCUS       AA376266      238 bp    mRNA            EST       21-APR-1997
DEFINITION  EST88915 HSC172 cells II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA376266
NID         g2028809
VERSION     AA376266.1  GI:2028809
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC191210
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .238
                      /organism="Homo sapiens"
                      /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):180944"
                      /db_xref="taxon:9606"
                      /clone_lib="HSC172 cells II"
                      /cell_type="fibroblast"
                      /cell_line="HSC172 (60PDL)"
                      /dev_stage="fetal"
BASE COUNT       48 a     70 c     50 g     66 t      4 others
ORIGIN

  Query Match           10.0%;  Score 23;  DB 12;  Length 238;
  Best Local Similarity  67.2%;  Pred. No. 3.12e-05;
  Matches    45;  Conservative    0;  Mismatches  22;  Indels   0;  Gaps   0;

Db     141 TGTTGCCTCCCTCCCAGGATCCCTTTGGTGAGTATGGTGTTCAGGATGCACCACCACCAC 200
           |||  ||||| |||||||| | |  |||  | ||||| ||   || ||| ||    || |
Qy      14 TGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTC 73

Db     201 CTCTAGA 207
           |||| ||
Qy      74 CTCTTGA 80


RESULT   10
LOCUS       AA323964      311 bp    mRNA            EST       20-APR-1997
DEFINITION  EST26816 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA323964
NID         g1976290
VERSION     AA323964.1  GI:1976290
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
```

```
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693635.

            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .311
                      /organism="Homo sapiens"
                      /note="Organ: brain; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
                      /db_xref="ATCC (inhost):124516"
                      /db_xref="taxon:9606"
                      /clone_lib="Cerebellum II"
                      /tissue_type="cerebellum"
                      /dev_stage="adult"
BASE COUNT       59 a    110 c     74 g     64 t      4 others
ORIGIN

  Query Match          10.0%;  Score 23;  DB 11;  Length 311;
  Best Local Similarity  89.7%;  Pred. No. 3.12e-05;
  Matches    26;  Conservative    0;  Mismatches   3;  Indels   0;  Gaps   0;

Db       8 TGGCCACTGGGGGGCCTGGGCTGCCCCTT 36
           |||||||| |||| ||||||||||||||
Qy      39 TGGCCACTATGGGGTCTGGGCTGCCCCTT 67


RESULT   11
LOCUS       W33870        364 bp    mRNA            EST       13-MAY-1996
DEFINITION  mc56b03.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:352493 5', mRNA sequence.
ACCESSION   W33870
NID         g1315794
VERSION     W33870.1  GI:1315794
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 18, 1995 this sequence version replaced gi:811194.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:224293
            Seq primer: ETPrimer
            High quality sequence stop: 360.
FEATURES              Location/Qualifiers
     source           1. .364
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                      T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                      14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                      State Univ., from 2 ]; double-stranded cDNA was ligated to
                      Eco RI adaptors (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of the modified
                      pT7T3 vector. Library went through one round of
                      normalization, and was constructed by Bento Soares and
                      M.Fatima Bonaldo.  "
                      /db_xref="taxon:10090"
                      /clone="IMAGE:352493"
                      /clone_lib="Soares mouse embryo NbME13.5 14.5"
                      /sex="unknown"
                      /tissue_type="embryo"
                      /dev_stage="13.5-14.5dpc total fetus"
                      /lab_host="DH10B"
BASE COUNT      100 a     73 c     87 g    104 t
ORIGIN

  Query Match          10.0%;  Score 23;  DB 34;  Length 364;
  Best Local Similarity  79.5%;  Pred. No. 3.12e-05;
  Matches    31;  Conservative    0;  Mismatches   8;  Indels   0;  Gaps   0;

Db     282 AACTGAAAGGGAAGGAGGCTTTTCCCACATATTTCTCCT 320
           |||||||   ||||||| ||||||  ||| ||| |||||
Qy     126 AACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCT 164


RESULT   12
LOCUS       AA769782      412 bp    mRNA            EST       29-DEC-1998
DEFINITION  ah71b05.s1 Soares_testis_NHT Homo sapiens cDNA clone 1321041 3',
            mRNA sequence.
ACCESSION   AA769782
NID         g2821020
VERSION     AA769782.1  GI:2821020
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 9, 1998 this sequence version replaced gi:936203.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 691   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 402.
FEATURES              Location/Qualifiers
     source           1. .412
                      /organism="Homo sapiens"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```

was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "
/db_xref="taxon:9606"
/clone="1321041"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"

BASE COUNT 111 a 78 c 116 g 107 t
ORIGIN

Query Match 10.0%; Score 23; DB 18; Length 412;
Best Local Similarity 67.2%; Pred. No. 3.12e-05;
Matches 45; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Db    262 TCTAGAGGTGGTGGTGGTGCATCCTGAACACCATACTCACCAAAGGGATCCTGGGAGGGA 321
          || ||||| ||    || ||| ||   || ||||| |  |||  | | ||||||||| |||
Cp     80 TCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATGGA 21

Db    322 GGCAACA 328
          ||  |||
Cp     20 GGAGACA 14
```

RESULT 13
LOCUS T15255 426 bp mRNA EST 28-JUL-1995
DEFINITION crs855 lambdaZAPST Ricinus communis cDNA clone pcrs855, mRNA sequence.
ACCESSION T15255
NID g688907
VERSION T15255.1 GI:688907
KEYWORDS EST.
SOURCE castor bean.
ORGANISM Ricinus communis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Euphorbiales; Euphorbiaceae; Ricinus.
REFERENCE 1 (bases 1 to 426)
AUTHORS vandeLoo,F.J., Turner,S. and Somerville,C.
TITLE Expressed sequence tags from developing castor seeds
JOURNAL Plant Physiol. 108, 1141-1150 (1995)
COMMENT
Contact: Somerville CR
Carnegie Institution
Carnegie Institution, 290 Panama St, Stanford, CA 94305
Tel: 4153251521
Email: crs@andrew.stanford.edu
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .426
/organism="Ricinus communis"
/strain="Baker 296"
/note="Vector: lambdaZAPII; Site_1: EcoR1; Site_2: XhoI; Poly(A)+ RNA was purified from developing stage III to stage V (Greenwood & Bewley, Can. J. Bot. 60:1751-1760, 1982) endosperm plus embryo of immature castor fruits. cDNA was synthesized and cloned into lambdaZAPII according to the instructions of the manufacturer (Stratagene): synthesis was primed from the poly(A) tail, and cloned directionally into XhoI (3') and EcoRI (5') sites. In few cases, sequence data indicated that this directionality was reversed. Partial cDNA clones predominate. "
/db_xref="taxon:3988"
/clone="pcrs855"
/clone_lib="lambdaZAPST"
BASE COUNT 126 a 61 c 99 g 129 t 11 others
ORIGIN

Query Match 10.0%; Score 23; DB 8; Length 426;
Best Local Similarity 71.4%; Pred. No. 3.12e-05;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
Db     77 TGACCATTCAAGTGAAGCTGAAGATGACTTGTCTGANTGATTGGNCCTA 125
          ||||  | ||| ||||||||||  | ||| |||||   |||   ||||
Qy    117 TGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTA 165
```

RESULT 14
LOCUS AA498488 481 bp mRNA EST 01-JUL-1997
DEFINITION vh44b06.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA clone IMAGE:889811 5' similar to gb:M30514 Mouse muscle nicotinic acetylcholine receptor gamma-subunit (MOUSE);, mRNA sequence.
ACCESSION AA498488
NID g2233511
VERSION AA498488.1 GI:2233511
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 481)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405167.

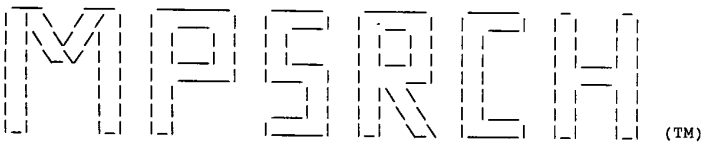
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:517771
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 461.
FEATURES Location/Qualifiers
source 1. .481
/organism="Mus musculus"
/strain="FVB/N"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTTGGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."
/db_xref="taxon:10090"
/clone="IMAGE:889811"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
BASE COUNT 120 a 131 c 124 g 106 t
ORIGIN

Query Match 10.0%; Score 23; DB 14; Length 481;
Best Local Similarity 76.7%; Pred. No. 3.12e-05;
Matches 33; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Db    277 CTTCAGTGTCAACAACCTGCTTGGAGCCCAGTGGCCAATATGG 319
          ||||||||||  |  | | |  ||||| |||||||||| |||||
```

```
Qy      8 CTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGG 50


RESULT   15
LOCUS       AF038250     1287 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF038250 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon9, mRNA sequence.
ACCESSION   AF038250
NID         g2815880
VERSION     AF038250.1  GI:2815880
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1287)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045085.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .1287
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon9"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      349 a    219 c    293 g    361 t     65 others
ORIGIN

  Query Match           10.0%;  Score 23;  DB 20;  Length 1287;
  Best Local Similarity  25.8%;  Pred. No. 3.12e-05;
  Matches    24;  Conservative   32;  Mismatches  37;  Indels   0;  Gaps   0;

Db    381 VVVATCNAGSCCVACVTCBCDCBTGGGDTBSHBVCMCBCBANADGATBATCGKVGGVBGA 440
          ::: ||  |:| : |:  : :|:|    |::::::| |: :   :| |:|   ::  : |
Qy     49 GGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 108

Db    441 STCMTHYNCCDCKTCGSAGVTVNHHDWSMAGGA 473
          : |   :   :|:| |:| :|: :  :::||||
Qy    109 GCCGGGTATGACTTTGCAACTGAAGCTGAAGGA 141


Search completed: Sun Oct 24 17:07:02 1999
Job time : 455 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 16:22:19 1999; MasPar time 475.19 Seconds
1394.147 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-1
Description: (1-239) from US09092296.seq
Perfect Score: 238
N.A. Sequence: 1 GGCCACCGGGACTTCAGTGT.....CCCTTCAGGGACCAGCGTCA 239
Comp: CCGGTGGCCCTGAAGTCACA.....GGGAAGTCCCTGGTCGCAGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl58
1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi
Database: genbank111
17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 9.607; Variance 4.824; scale 1.991

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	112	47.1	47323	31	AC005937	Homo sapiens clone UWG	7.72e-65
	2	47	19.7	7218	25	I66494	Sequence 14 from paten	2.10e-15
c	3	36	15.1	215	25	I28278	Sequence 5 from patent	5.87e-08
	4	33	13.9	216021	31	HUAC004787	Homo sapiens Chromosom	4.91e-06
	5	31	13.0	965	25	AR024229	Sequence 22 from paten	8.69e-05
	6	29	12.2	74371	31	AC005369	Homo sapiens chromosom	1.43e-03
	7	28	11.8	1056	23	MVU87256	Mustela vison GT dinuc	5.63e-03
c	8	28	11.8	216021	31	HUAC004787	Homo sapiens Chromosom	5.63e-03
	9	27	11.3	215	25	I28278	Sequence 5 from patent	2.17e-02
c	10	26	10.9	965	25	AR024229	Sequence 22 from paten	8.16e-02
	11	25	10.5	565	25	E04076	gDNA encoding envelope	3.00e-01
	12	25	10.5	60966	31	AC003030	Homo sapiens chromosom	3.00e-01
c	13	24	10.1	60	25	A62989	Sequence 1 from Patent	1.07e+00
c	14	24	10.1	1056	23	MVU87256	Mustela vison GT dinuc	1.07e+00
c	15	24	10.1	175793	31	AC005920	Homo sapiens chromosom	1.07e+00
	16	23	9.7	3290	29	HUMPFUCAS	H.sapiens fucosidase p	3.72e+00
	17	23	9.7	4472	17	NMSERBORF	Neisseria meningitidis	3.72e+00
c	18	23	9.7	133457	30	AC003999	Human PAC clone DJ1139	3.72e+00
	19	23	9.7	192853	19	AC005096	Homo sapiens clone RG3	3.72e+00
	20	23	9.7	203418	19	AC004947	Homo sapiens clone DJ1	3.72e+00
c	21	22	9.2	30	25	A62994	Sequence 6 from Patent	1.25e+01
c	22	22	9.2	108	21	D87227	Trypanosoma cruzi mRNA	1.25e+01
c	23	22	9.2	1738	29	HUMGOLGINB	Human (clone SY2/10) g	1.25e+01
c	24	22	9.2	2180	21	DROSYT	D.melanogaster synapto	1.25e+01
	25	22	9.2	2273	32	MMBAKEXN	Mus musculus Bak gene,	1.25e+01
c	26	22	9.2	3088	25	AR014574	Sequence 1 from patent	1.25e+01
c	27	22	9.2	3088	11	E10775	DNA encoding part of A	1.25e+01
c	28	22	9.2	3628	29	AB018292	Homo sapiens mRNA for	1.25e+01
c	29	22	9.2	3890	30	HSA010901	Homo sapiens MUC4 gene	1.25e+01
	30	22	9.2	3914	32	AF090866	Mus musculus CDO (Cdo)	1.25e+01
c	31	22	9.2	6640	29	D63997	Homo sapiens mRNA for	1.25e+01
	32	22	9.2	67919	20	AC006542	WORKING DRAFT SEQUENCE	1.25e+01
	33	22	9.2	82098	31	AC006252	Homo sapiens 3p21.1 co	1.25e+01
c	34	22	9.2	124700	22	AC005558	Drosophila melanogaste	1.25e+01
	35	22	9.2	135039	31	AC006060	Homo sapiens 3p22-8 PA	1.25e+01
	36	22	9.2	151187	19	HS460J8	Human DNA sequence ***	1.25e+01
c	37	22	9.2	175339	31	AC005772	Homo sapiens chromosom	1.25e+01
c	38	22	9.2	241407	20	AC003059	Mouse Chromosome 10 BA	1.25e+01
	39	21	8.8	1755	29	HSZ97017	Homo sapiens mRNA for	4.06e+01
	40	21	8.8	2817	27	AB012632	Robinia pseudoacacia g	4.06e+01
	41	21	8.8	3043	32	AB010141S4	Mus musculus gene for	4.06e+01
	42	21	8.8	3247	32	AB010140	Mus musculus mRNA for	4.06e+01
	43	21	8.8	46213	28	SPBC18H10	S.pombe chromosome II	4.06e+01
	44	21	8.8	113887	20	AC006158	Homo sapiens clone NH0	4.06e+01
	45	21	8.8	130560	19	HS508I15	Human DNA sequence ***	4.06e+01

ALIGNMENTS

RESULT 1
LOCUS AC005937 47323 bp DNA PRI 05-NOV-1998
DEFINITION Homo sapiens clone UWGC:370M23.002 from 6p21, complete sequence.
ACCESSION AC005937
NID g3845393
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 47323)
AUTHORS Janer,M., Guillaudeux,T., Vu,Q., Kutyavin,T., Harter,H. and Geraghty,D.E.
TITLE Large scale sequence analysis of the human MHC class I region
JOURNAL Unpublished (1998)
REMARK Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024
REFERENCE 2 (bases 1 to 47323)
AUTHORS Geraghty,D.E. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhcrc.org)
COMMENT Overlapping Sequences:
5': UWGC:370m23.013 (Genbank Accession: AC005530)
3': UWGC:y67c112 (Genbank Accession: AC004211)

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than

1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 75.5%
DS or two chemistry coverage: 98.9%
Single stranded regions: 3

Sequence Validation:
This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BglII		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
1069.11	1050.00	889.55	866.00	30541.40	30653.00
20320.67	20855.00	1050.18	1015.00	3279.08	3231.00
2171.50	2147.00	7268.78	7196.00		
2560.20	2531.00	10085.80	9992.00		
4335.42	4269.00	11212.78	11131.00		
2698.62	2628.00				
1927.50	1887.00				
3130.46	3090.00				
2166.69	2129.00				
2044.67	2005.00				

```
FEATURES             Location/Qualifiers
     source          1. .47323
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21"
                     /sub_clone="UWGC:370M23.002"
                     /clone_lib="Research Genetics BAC Library"
     repeat_region   3647. .3932
                     /rpt_family="Alu"
     repeat_region   complement(4999. .5277)
                     /rpt_family="Alu"
     repeat_region   6285. .6572
                     /rpt_family="Alu"
     repeat_region   complement(6972. .7050)
                     /rpt_family="MLT1"
     repeat_region   7286. .7584
                     /rpt_family="Alu"
     repeat_region   complement(8164. .8609)
                     /rpt_family="Alu"
     repeat_region   complement(21287. .21895)
                     /rpt_family="Alu"
     repeat_region   22715. .22957
                     /rpt_family="Alu"
     repeat_region   25510. .25802
                     /rpt_family="Alu"
     repeat_region   27835. .28010
                     /rpt_family="MER20"
     repeat_region   31295. .31594
                     /rpt_family="Alu"
     repeat_region   33515. .33767
                     /rpt_family="Alu"
     repeat_region   34223. .34290
                     /rpt_family="MIR"
     repeat_region   37372. .37648
                     /rpt_family="Alu"
     repeat_region   38526. .38700
                     /rpt_family="MER3"
     repeat_region   39583. .40010
                     /rpt_family="Alu"
     repeat_region   40046. .40156
                     /rpt_family="Alu"
     repeat_region   43194. .43372
                     /rpt_family="MER5"
     variation       43325
                     /note="clonal variation with 3' overlapping clone"
     variation       44149
                     /note="clonal variation with 3' overlapping clone"
     variation       44451
                     /note="clonal variation with 3' overlapping clone"
     variation       44537
                     /note="clonal variation with 3' overlapping clone"
     variation       44814
                     /note="clonal variation with 3' overlapping clone"
     variation       44965
                     /note="clonal variation with 3' overlapping clone"
     variation       45760
                     /note="clonal variation with 3' overlapping clone"
     variation       45900
                     /note="clonal variation with 3' overlapping clone"
     variation       46851
                     /note="clonal variation with 3' overlapping clone"
     variation       46859
                     /note="clonal variation with 3' overlapping clone"
     variation       47032
                     /note="clonal variation with 3' overlapping clone"
     variation       47240. .47256
                     /note="clonal variation with 3' overlapping clone -
                     insertion of 17bp repeat"
BASE COUNT    11556 a  11489 c  12284 g  11994 t
ORIGIN

  Query Match           47.1%;  Score 112;  DB 31;  Length 47323;
  Best Local Similarity 100.0%;  Pred. No. 7.72e-65;
  Matches   112;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db  35117 GGCCACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTG 35176
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GGCCACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTG 60

Db  35177 GGCTGCCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 35228
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GGCTGCCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 112


RESULT    2
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

```
  Query Match           19.7%;  Score 47;  DB 25;  Length 7218;
  Best Local Similarity  2.3%;  Pred. No. 2.10e-15;
  Matches     5;  Conservative  129;  Mismatches  88;  Indels   0;  Gaps   0;

Db   1219 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1278
          ::::  : :::::::: ::::     :  :   :  :: :: :    :::   :: ::::::
Qy     12 CTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTT 71

Db   1279 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1338
           :::::::::: :::::::: :  ::: : :    :    ::   : :  :::: :
Qy     72 GTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAA 131

Db   1339 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1398
          ::    ::      :::::::: :   :::::::: :   :::  ::::::    ::
Qy    132 CTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTG 191

Db   1399 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440
          :::      ::: :::::::::: :::::::::  | ||||
Qy    192 CTTGAAAANTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCA 233


RESULT    3
LOCUS       I28278        215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
VERSION     I28278.1  GI:1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="unknown"
BASE COUNT       15 a       8 c      25 g      26 t     141 others
ORIGIN

  Query Match           15.1%;  Score 36;  DB 25;  Length 215;
  Best Local Similarity  14.7%;  Pred. No. 5.87e-08;
  Matches    28;  Conservative   77;  Mismatches  83;  Indels   2;  Gaps   2;

Db      1 MTNVTMSSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNK 60
          :| : :::  :: : |:  ::|:::|     : |    : :|:|:  ||:::  :: : :
Cp    189 ATTCCAGGAAGCTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTC-AGT 131

Db     61 YSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTD-GNRSGADS 119
          :::|   :    :   : ::: :| :   ::  :|::: |:| :::    : | ::||
Cp    130 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 71

Db    120 YGSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTKKHAKNSADGKVGSKNNGDRNN 179
           |::: |  :  :       |:::: :   ::|::  |:: |: :|:| : :   |::
Cp     70 AGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATGGAGGAGACACTGAAGT 11

Db    180 RYGTGTKSNV 189
          :  ||:: :
Cp     10 CCCGGTGGCC 1


RESULT    4
LOCUS       HUAC004787 216021 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
NID         g3337381
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .216021
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="#16q21-22"
                     /clone="A-952F10"
     STS             27765. .27872
                     /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02280"
     STS             73826. .73943
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     STS             175801. .175945
                     /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09703"
     STS             175810. .175945
                     /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09935"
     STS             199463. .199572
                     /note="9824, WI-3555, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G04338"
BASE COUNT    60960 a   51778 c   49172 g   53987 t    124 others
ORIGIN

  Query Match           13.9%;  Score 33;  DB 31;  Length 216021;
  Best Local Similarity  12.2%;  Pred. No. 4.91e-06;
  Matches    18;  Conservative   72;  Mismatches  57;  Indels   0;  Gaps   0;

Db   1436 WYCSSYTYCCYYCSRARRRGKKKYMWKSMRRARRRGSAGKKKKKYYYYYCYYYYYYYYCM 1495
          ::| :: :||:  ::  :  ::: :  : :  ::: : |:: :::::::  :::::::|
Qy     22 TCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCT 81

Db   1496 GRAMMAAAWYYKRRSCMAWYYYMRGRARKTYYYMARRRGCARSYKKYMAMYWAAWTTWCA 1555
           :|::    ::::   :   :  :|:|     :: ::    ::::: :|::: |: : |
Qy     82 TGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTGA 141

Db   1556 AAAMAAMYTTYYWMMWYYKKWMYTCCT 1582
           |  :  ::||::: :    ::::||||
```

```
Qy    142 AGGAGTCTTTTCTGACAAATTCCTCCT 168


RESULT    5
LOCUS       AR024229      965 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 22 from patent US 5795961.
ACCESSION   AR024229
NID         g3977523
VERSION     AR024229.1  GI:3977523
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and
            Kitamura,K.
  TITLE     Recombinant human anti-Lewis b antibodies
  JOURNAL   Patent: US 5795961-A 22 18-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .965
                     /organism="unknown"
BASE COUNT      192 a    170 c    226 g    205 t    172 others
ORIGIN

  Query Match           13.0%;  Score 31;  DB 25;  Length 965;
  Best Local Similarity  20.2%;  Pred. No. 8.69e-05;
  Matches    24;  Conservative   52;  Mismatches  42;  Indels   1;  Gaps   1;

Db    845 SVTAADTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHTVKDMTSSSSASVGDRVTTCRSST 904
          ::|  :| ::  ::| :::::|| ::   |: :: :|::::|::::   |: :| | : |
Qy     23 CCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTT 82

Db    905 THGNGNTYYWYKGKAKYRVSNRSGVSRSGSGSGTDYTTSSDATYYCGTHARTGGTKVKG 963
          :    |:::::| |:: :  : |: :: :|:  :  |  :| :: | :| ||  :::|
Qy     83 GACC-CTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG 140


RESULT    6
LOCUS       AC005369    74371 bp    DNA             PRI       01-AUG-1998
DEFINITION  Homo sapiens chromosome 5, BAC clone 119j3 (LBNL H175), complete
            sequence.
ACCESSION   AC005369
NID         g3367505
VERSION     AC005369.1  GI:3367505
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 74371)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 74371)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 74371)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .74371
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="5q"
                     /clone="119j3"
                     /chromosome="5"
                     /note="LBNL H175"
     repeat_region   893. .1030
                     /rpt_family="Alu"
     repeat_region   2295. .2438
                     /rpt_family="Alu"
     repeat_region   2818. .2859
                     /note="(GT)21"
                     /rpt_type=tandem
                     /rpt_unit=GT
     misc_feature    join(3246. .3410,3721. .3828)
                     /standard_name="RLF"
                     /note="65% & 69% protein identity GenPept:U22377"
     repeat_region   3431. .3724
                     /rpt_family="Alu"
     repeat_region   3707. .3728
                     /note="(A)22"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   4366. .4661
                     /rpt_family="Alu"
     repeat_region   5327. .5602
                     /rpt_family="Alu"
     repeat_region   6586. .6956
                     /rpt_family="L1"
     repeat_region   6647. .6684
                     /note="(CA)19"
                     /rpt_type=tandem
                     /rpt_unit=CA
     repeat_region   7113. .7373
                     /rpt_family="Alu"
     repeat_region   complement(7830. .8185)
                     /standard_name="possible repeat"
     repeat_region   8258. .8503
                     /rpt_family="Alu"
     repeat_region   9070. .9387
                     /rpt_family="Alu"
     repeat_region   complement(9740. .9845)
                     /rpt_family="MER42"
     repeat_region   complement(10440. .11015)
                     /rpt_family="Alu"
     repeat_region   11950. .12250
                     /rpt_family="Alu"
     repeat_region   12057. .12085
                     /note="(A)29"
                     /rpt_type=tandem
                     /rpt_unit=A
     repeat_region   12365. .12645
                     /rpt_family="Alu"
     repeat_region   13727. .13750
                     /note="(AC)12"
                     /rpt_type=tandem
                     /rpt_unit=AC
     repeat_region   13783. .14024
                     /rpt_family="L1"
     repeat_region   14175. .14470
                     /rpt_family="Alu"
     repeat_region   complement(14906. .15259)
                     /standard_name="possible repeat"
     repeat_region   15300. .15613
                     /rpt_family="Alu"
     repeat_region   16671. .16690
                     /note="(T)20"
                     /rpt_type=tandem
                     /rpt_unit=T
     repeat_region   complement(16675. .16977)
                     /rpt_family="Alu"
     repeat_region   complement(16993. .17085)
                     /rpt_family="MER42"
     repeat_region   complement(17678. .18276)
```

```
                    /rpt_family="Alu"
     repeat_region  19305. .19583
                    /rpt_family="Alu"
     repeat_region  19914. .19945
                    /note="(GTTT)8"
                    /rpt_type=tandem
                    /rpt_unit=GTTT
     repeat_region  complement(19943. .20222)
                    /rpt_family="Alu"
     repeat_region  21202. .21496
                    /rpt_family="Alu"
     repeat_region  21736. .22035
                    /rpt_family="Alu"
     repeat_region  22017. .22038
                    /note="(A)22"
                    /rpt_type=tandem
                    /rpt_unit=A
     repeat_region  22321. .22457
                    /rpt_family="Alu"
     repeat_region  complement(22638. .22981)
                    /standard_name="possible repeat"
     repeat_region  23022. .23326
                    /rpt_family="Alu"
     repeat_region  23473. .23761
                    /rpt_family="Alu"
     repeat_region  23744. .23767
                    /note="(A)24"
                    /rpt_type=tandem
                    /rpt_unit=A
     repeat_region  complement(24639. .24694)
                    /rpt_family="MER42"
     repeat_region  complement(25349. .25713)
                    /rpt_family="Alu"
     repeat_region  complement(25727. .26471)
                    /rpt_family="Alu"
     repeat_region  27191. .27477
                    /rpt_family="Alu"
     repeat_region  27774. .28057
                    /rpt_family="Alu"
     repeat_region  28040. .28066
                    /note="(A)27"
                    /rpt_type=tandem
                    /rpt_unit=A
     repeat_region  28712. .28930
                    /rpt_family="MER20"
     misc_feature   complement(28769. .28838)
                    /note="GRAIL 2 excellent exon, frame 0"
     repeat_region  28987. .29214
                    /rpt_family="Alu"
     misc_feature   join(29485. .29559,29598. .29779)
                    /note="95% identity dbSTS:G14522 (SHGC-11312)"
     misc_feature   29495. .29976
                    /standard_name="AI027942"
                    /note="100% identity EST ov84a10.x1"
     misc_feature   complement(30401. .30536)
                    /note="GRAIL 2 excellent exon, frame 2"
     repeat_region  complement(30682. .30733)
                    /rpt_family="MIR"
     misc_feature   complement(31573. .31724)
                    /note="GRAIL 2 excellent exon, frame 1"
     misc_feature   complement(32159. .32232)
                    /note="GRAIL 2 excellent exon, frame 2"
     repeat_region  32388. .32488
                    /rpt_family="MLT1"
     repeat_region  32617. .32908
                    /rpt_family="Alu"
     repeat_region  32977. .33088
                    /rpt_family="MLT1"
     repeat_region  complement(33670. .33785)
                    /rpt_family="Alu"
     misc_feature   complement(34021. .34144)
                    /note="GRAIL 2 excellent exon, frame 2"
     misc_feature   complement(35238. .35331)
                    /note="GRAIL 2 excellent exon, frame 1"
     repeat_region  36392. .36663
                    /rpt_family="Alu"
     misc_feature   36901. .37222
                    /note="100% identity EST ou55c09.x1"
                    /db_xref="dbEST:AI025011"
     STS            36901. .37164
                    /standard_name="STSG-9983"
                    /db_xref="dbSTS:G26554"
     misc_feature   complement(37595. .37654)
                    /note="GRAIL 2 excellent exon, frame 0"
     misc_feature   complement(join(38063. .38218,38462. .38578,38741. .38995,
                    39071. .39205,39532. .39630,39935. .40048,40300. .40410,
                    40503. .40661,41868. .41972,42103. .42225,42492. .42569,
                    44379. .44507))
                    /standard_name="histidyl-tRNA synthetase"
                    /note="78%-100% protein identity GenPept:U18937"
     misc_feature   complement(38069. .38215)
                    /note="GRAIL 2 excellent exon, frame 0"
...
Note: remainder of annotations omitted.

  Query Match           12.2%;  Score 29;  DB 31;  Length 74371;
  Best Local Similarity  16.7%;  Pred. No. 1.43e-03;
  Matches    14;  Conservative   45;  Mismatches  24;  Indels   1;  Gaps   1;

Db  15929 CMSRSKSKRGWGYRSWKKYRCAMWMTCKSS-KCWCWSYRMRMKCYSCSYCYCSSGKKYWC 15987
          |:::: ::  :|::  : : | ::: | :: :| |::::  : |:  :    : |::::
Qy      6 CCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTG 65

Db  15988 RCSMYWYTCYYSYKYYWSMSYCTC 16011
           |:::: ||:::::::::::::|||
Qy     66 CCCCTTGTCCTCCTCTTGACCCTC 89


RESULT    7
LOCUS       MVU87256     1056 bp    DNA             MAM       02-JAN-1999
DEFINITION  Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION   U87256
NID         g4099442
VERSION     U87256.1  GI:4099442
KEYWORDS    .
SOURCE      American mink.
  ORGANISM  Mustela vison
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
            Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /chromosome="1"
                     /map="1q"
                     /note="primers: 1167F: agcccctgcatatctacttctt, 1167R:
                     gaggatcttaccgctgttgag"
     primer_bind     98. .119
                     /standard_name="1167F"
     primer_bind     complement(300. .320)
                     /standard_name="1167R"
BASE COUNT      211 a    221 c    210 g    225 t    189 others
ORIGIN

  Query Match           11.8%;  Score 28;  DB 23;  Length 1056;
  Best Local Similarity  12.8%;  Pred. No. 5.63e-03;
  Matches    12;  Conservative   49;  Mismatches  33;  Indels   0;  Gaps   0;

Db    576 SCRHDVBMWSKWCWGKKSCCKSTGKDKMSGCAYCGKRMRRCSRYMMMRMCVGSGTWARCC 635
          :|:: :: ::::|:|:: |  :|  ::::     | :::: :::::: : :|:||  :
Qy      3 CCACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGG 62
```

```
Db    636 CCDKKSGSGDKSHCKSKRKYKWMDRYHBCKSMCA 669
          | :  :   ::::  :: :::::  ::: :: ||
Qy     63 CTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCA 96


RESULT    8
LOCUS       HUAC004787 216021 bp    DNA             PRI       24-JUL-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
NID         g3337381
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES              Location/Qualifiers
     source           1. .216021
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="16"
                      /map="#16q21-22"
                      /clone="A-952F10"
     STS              27765. .27872
                      /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                      sapiens"
                      /db_xref="dbSTS:G02280"
     STS              73826. .73943
                      /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                      sapiens"
                      /db_xref="dbSTS:G02122"
     STS              175801. .175945
                      /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                      /db_xref="dbSTS:G09703"
     STS              175810. .175945
                      /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                      /db_xref="dbSTS:G09935"
     STS              199463. .199572
                      /note="9824, WI-3555, Chr. 16, Homo sapiens"
                      /db_xref="dbSTS:G04338"
BASE COUNT    60960 a  51778 c  49172 g  53987 t    124 others
ORIGIN

  Query Match           11.8%;  Score 28;  DB 31;  Length 216021;
  Best Local Similarity  14.7%;  Pred. No. 5.63e-03;
  Matches    19;  Conservative   59;  Mismatches  51;  Indels   0;  Gaps   0;

Db   1457 KKYMWKSMRRARRRGSAGKKKKKYYYYYCYYYYYYYYCMGRAMMAAAWYYKRRSCMAWYY 1516
          :: : ::::: :::|:||: ::     : :::   :      :  || :::::: :|:
Cp    226 TGAAGGGAGATGGAGGAGGAGGCAGANTTTTCAAGCAATTCCAGGAAGCTGGACTCATAG 167

Db   1517 YMRGRARKTYYYMARRRGCARSYKKYMAMYWAAWTTWCAAAAMAAMYTTYYWMMWYYKKW 1576
           ::|:| :| ::: :::  |  : :::| ::  : |:  |||   ::  ::  : :::::
Cp    166 GAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTT 107

Db   1577 MYTCCTGAG 1585
           ::   ||||
Cp    106 CCATGTGAG 98


RESULT    9
LOCUS       I28278        215 bp    DNA             PAT       30-OCT-1996
DEFINITION  Sequence 5 from patent US 5569830.
ACCESSION   I28278
NID         g1819054
VERSION     I28278.1  GI:1819054
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.
  TITLE     Plant inhibitors of fungal polygalacturonases and their use to
            control fungal disease
  JOURNAL   Patent: US 5569830-A 5 29-OCT-1996;
FEATURES              Location/Qualifiers
     source           1. .215
                      /organism="unknown"
BASE COUNT       15 a      8 c     25 g     26 t    141 others
ORIGIN

  Query Match           11.3%;  Score 27;  DB 25;  Length 215;
  Best Local Similarity  20.0%;  Pred. No. 2.17e-02;
  Matches    22;  Conservative   42;  Mismatches  45;  Indels   1;  Gaps   1;

Db     18 CNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAK 77
          | :: : :| |  ::| :   :  |:|: :| ::: : |:  : :|   :|   :|| :
Qy     89 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG-AAGGAGT 147

Db     78 THYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
           :::|:| :: |: : :| ::  :   :: | |:   :||: :|: : |:
Qy    148 CTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAA 197


RESULT   10
LOCUS       AR024229      965 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 22 from patent US 5795961.
ACCESSION   AR024229
NID         g3977523
VERSION     AR024229.1  GI:3977523
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and
            Kitamura,K.
  TITLE     Recombinant human anti-Lewis b antibodies
  JOURNAL   Patent: US 5795961-A 22 18-AUG-1998;
FEATURES              Location/Qualifiers
     source           1. .965
                      /organism="unknown"
BASE COUNT      192 a    170 c    226 g    205 t    172 others
ORIGIN
```

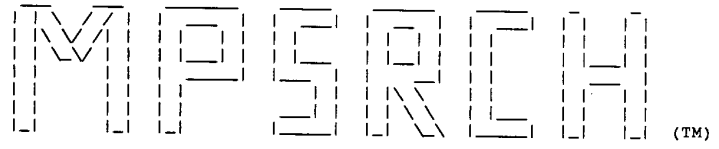
```
  Query Match           10.9%;  Score 26;  DB 25;  Length 965;
  Best Local Similarity  20.0%;  Pred. No. 8.16e-02;
  Matches    18;  Conservative   41;  Mismatches  30;  Indels   1;  Gaps   1;

Db    781 GTRHTVHVSGGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKRGRVTMADTSS 840
          || :|::::||: :|:| | ::: |:   ::  :  |: || :    |::|:: :|: ::
Cp    124 GTCATACCCGGCCCTGT-TCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 66

Db    841 NSRSSVTAADTAVYYCVRGRSYDSDGGDYW 870
            ::::  |   : : : |:::::| :::
Cp     65 CAGCCCAGACCCCATAGTGGCCACTGCGCT 36


RESULT   11
LOCUS       E04076        565 bp    RNA             PAT       29-SEP-1997
DEFINITION  gDNA encoding envelope region of type C hepatitis virus.
ACCESSION   E04076
NID         g2172286
VERSION     E04076.1  GI:2172286
KEYWORDS    JP 1992349885-A/1.
SOURCE      Hepatitis C virus.
  ORGANISM  Hepatitis C virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
            Hepatitis C-like viruses.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Morinaga,T., Chayama,K., Kumada,H. and Ichikawa,Y. .
  TITLE     NUCLEIC ACID FRAGMENT WITHIN ENVELOPE REGION OF HEPATITIS C VIRUS
            AND METHOD FOR DETECTING THE SAME
  JOURNAL   Patent: JP 1992349885-A 1 04-DEC-1992;
            TEIJIN LTD
COMMENT     OS   Hepatitis C virus
            PN   JP 1992349885-A/1
            PD   04-DEC-1992
            PF   29-MAY-1991 JP 1991152169
            PI   MORINAGA TSUTAE, CHAYAMA KAZUAKI, KUMADA HIROMITSU, PI
            ICHIKAWA YATARO
            PC   C12N15/10,C12Q1/68,C12Q1/70//C12N15/11;
            CC   strandedness: Single;
            CC   topology: Linear;.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Hepatitis C virus"
                     /db_xref="taxon:11103"
BASE COUNT       60 a     93 c    107 g     85 t    220 others
ORIGIN

  Query Match           10.5%;  Score 25;  DB 25;  Length 565;
  Best Local Similarity  34.1%;  Pred. No. 3.00e-01;
  Matches    30;  Conservative   29;  Mismatches  28;  Indels   1;  Gaps   1;

Db    400 GGRTCYGTYTTYCTYRTYTCYCABYTGTTTACCYTCTMBMCYMGSMDGYAYGDGACRDYR 459
          ||:||:|   | | : |:  :| ::| || |||:||  :  ::|:   :|:|: ||:: :
Qy     54 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 113

Db    460 CARGRYTGYAAYTGYTCDMTYTAYSCYG 487
          | :|::  : |:| : |:: : | :|:|
Qy    114 CCGGGTA-TGACTTTGCAACTGAAGCTG 140


RESULT   12
LOCUS       AC003030    60966 bp    DNA             PRI       06-JAN-1999
DEFINITION  Homo sapiens chromosome 19, overlapping cosmids R29828 and F25496,
            complete sequence.
ACCESSION   AC003030
NID         g4092821
VERSION     AC003030.1  GI:4092821
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
            Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S.,
            Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 2 Mb contig in 19p12 between UBA52 and
            D19S455
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1997) Human Genome Center, Joint Genome
            Institute/ Lawrence Livermore National Laboratory, 7000 East Ave.,
            Livermore, CA 94551, USA
REFERENCE   3  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE   4  (bases 1 to 60966)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     On Jan 4, 1999 this sequence version replaced gi:2570224.
            Map and sequence oriented from p telomere to centromere. This
            accession is comprised of overlapping cosmids R29828 (bases 1 to
            40,974) and F25496 (bases 23,336 to 60,966). R29828 is separated
            from cosmid F23858 (AC004475) to the left by a sequence gap of
            approximately 14 kb, which is to be filled by sequencing a
            restriction fragment from cosmid R27236 (currently in progress).
            Cosmid F25496 is separated from cosmid R31863 to the right by an
            expected sequence gap of at least 10 kb. Additional chr 19 map and
            sequence information are available at:
            http://www-bio.llnl.gov/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1. .60966
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="R29828-F25496"
                     /chromosome="19"
                     /map="19p12 between UBA52 and D19S455."
                     /cell_line="UV5HL9-5B for F25496, and 5HL2-B for R29828"
                     /clone_lib="LL19NC02 (for F25496) and LL19NC03 (for
                     R29828) chromosome 19-specific cosmid libraries"
                     /note="Cosmid libraries constructed at LLNL from
                     flow-sorted chromosomes from human-hamster hybrid cell
                     lines UV5HL9-5B and 5HL2-B, each of which carries
                     chromosome 19 as its only human chromosome."
     repeat_region   18. .219
                     /rpt_family="AluJo"
     repeat_region   530. .824
                     /rpt_family="AluY"
     repeat_region   930. .1066
                     /rpt_family="FLAM_C"
     misc_feature    complement(1148. .1315)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: good, score: 51.000"
     repeat_region   complement(1477. .1706)
                     /rpt_family="AluJo"
     repeat_region   2105. .2406
                     /rpt_family="AluSp"
     repeat_region   2513. .2809
                     /rpt_family="AluJo"
     misc_feature    5028. .5045
                     /note="DDS similarity to overlapping ESTs:~AA963316
                     UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                     clone UI-R-E1- gi-c-06-0-UI 3'; (220. .237); 100%
                     identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                     Soares Rattus sp. cDNA clone RKIBE38 3' end; (240. .257);
```

```
                100% identity.~"
     CDS        join(<5028. .5045,5432. .5497,6480. .6575,8155. .8249,
                10156. .10183,10562. .10717,11845. .11991,12086. .12176,
                13147. .13250,14614. .14679,18199. .18248,18465. .18542,
                18635. .18746,23665. .23755,24574. .24709,37283. .37493)
                /note="Hypothetical human protein (partial)"
                /codon_start=1
                /evidence=not_experimental
                /product="R29828_1"
                /protein_id="AAD03161.1"
                /db_xref="PID:g4106983"
                /db_xref="GI:4106983"
                /translation="WLISQQIPQFEDVKFEAASLLSELYCQENSVDAAKPLLRKAIQI
                SQQTPYWHCRLLFQLAQLHTLEKDLVSACDLLGVGAEYARVVGSEYTRALFLLSKGML
                LLMERKLQEVHPLLTLCGQIVENWQGNPIQKESLRVFFLVLQVTHYLDAGQVKSVKPC
                LKQLQQCIQTISTLHDDEILPSNPADLFHWLPKEHMCVLVYLVTVMHSMQAGYLEKAQ
                KYTDKALMQLEKLKMLDCSPILSSFQVILLEHIIMCRLVTGHKATALQEGLYCVSVNC
                MDNAEAQFTTALRLYSLLERINPDHSFPVSSHCLRAAAFYVRGLFSFFQGRYNEAKRF
                LRETLKMSNAEDLNRLTACSLVLLGHIFYVLGNHRESNNMVVPAMQLASKIPDMSVQL
                WSSALLRDLNKACGNAMDAHEAAQMHQNFSQQLLQDHIEACSLPEHNLITVRVGHWEG
                IAATRLPVAFMEGQSTTTLYSADIWKHIQEDTEGSESAWAPVFRCEIKSLRPQTRAII
                LLPPLD"
     misc_feature   complement(5186. .5264)
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 86.000"
     misc_feature   5432. .5497
                /note="DDS similarity to overlapping ESts:~AA963316
                UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                clone UI-R-E1- gi-c-06-0-UI 3'; (238. .303); 88%
                identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                Soares Rattus sp. cDNA clone RKIBE38 3' end; (258. .323);
                88% identity."
     repeat_region  complement(5602. .5883)
                /rpt_family="AluJo"
     misc_feature   6480. .6575
                /note="DDS similarity to overlapping ESTs:~AA963316
                UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                clone UI-R-E1- gi-c-06-0-UI 3'; (304. .399); 89%
                identity.~~AA893275 EST197078 Normalized rat kidney, Bento
                Soares Rattus sp. cDNA clone RKIBE38 3' end; (324. .419);
                89% identity.~~(6480. .6660) predicted exon, program:
                grail2exons_human_1.3, frame: 2, quality: excellent,
                score: 76.000"
     misc_feature   8155. .8249
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 94.000~~DDS
                similarity to overlapping ESTs:~(8155. .8237) AA963316
                UI-R-E1-gi-c-06-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA
                clone UI-R-E1- gi-c-06-0-UI 3'; (400. .482); 88%
                identity.~~(8155. .8240) AA893275 EST197078 Normalized rat
                kidney, Bento Soares Rattus sp. cDNA clone RKIBE38 3' end;
                (420. .505); 85% identity."
     repeat_region  8351. .8477
                /rpt_family="AluJb"
     repeat_region  8480. .8768
                /rpt_family="AluSc"
     misc_feature   8778. .8837
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 83.000"
     repeat_region  complement(8872. .8996)
                /rpt_family="AluJb"
     repeat_region  9052. .9196
                /rpt_family="AluSx"
     repeat_region  9197. .9493
                /rpt_family="AluY"
     repeat_region  9494. .9658
                /rpt_family="AluSx"
     repeat_region  9668. .9967
                /rpt_family="AluY"
     misc_feature   10156. .10183
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 86.000"
     misc_feature   10562. .10721
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 100.000"
     misc_feature   11845. .11991
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 97.000"
     misc_feature   12086. .12176
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 1, quality: excellent, score: 96.000"
     misc_feature   13147. .13250
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 100.000"
     repeat_region  13440. .13741
                /rpt_family="AluSx"
     misc_feature   14614. .14679
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 100.000"
     misc_feature   16463. .16747
                /note="DDS similarity to AA212842 mw82d12.r1 Soares mouse
                NML Mus musculus cDNA clone 677207 5'; (200. .468); 80%
                identity."
     repeat_region  17395. .17687
                /rpt_family="AluJb"
     misc_feature   18199. .18248
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 89.000"
     misc_feature   18465. .18542
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 0, quality: excellent, score: 100.000"
     misc_feature   18635. .18746
                /note="predicted exon, program: grail2exons_human_1.3,
                frame: 2, quality: excellent, score: 97.000"
     repeat_region  complement(19286. .19309)
                /rpt_family="POLY_A"
     repeat_region  19323. .19623
                /rpt_family="AluSx"
     repeat_region  19632. .19928
                /rpt_family="AluSx"
     repeat_region  19929. .20226
                /rpt_family="AluY"
     repeat_region  20234. .20341
                /rpt_family="AluJo/FLAM"
     repeat_region  complement(20344. .20655)
...
Note: remainder of annotations omitted.

  Query Match          10.5%;  Score 25;  DB 31;  Length 60966;
  Best Local Similarity  85.7%;  Pred. No. 3.00e-01;
  Matches    30;  Conservative    0;  Mismatches   5;  Indels   0;  Gaps   0;

Db  13976 TGTCCACGATGGGATCTGGGCTGCCCTTTGCCCTC 14010
          || |||| ||||| |||||||||||| ||| ||||
Qy     43 TGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTC 77


RESULT   13
LOCUS       A62989         60 bp    DNA             PAT       12-MAR-1998
DEFINITION  Sequence 1 from Patent WO9720068.
ACCESSION   A62989
NID         g3716861
VERSION     A62989.1  GI:3716861
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 60)
  AUTHORS   Oerum,H. and Seeger,C.
  TITLE     METHOD FOR GENERATING MULTIPLE DOUBLE STRANDED NUCLEIC ACIDS
  JOURNAL   Patent: WO 9720068-A 1 05-JUN-1997;
            BOEHRINGER MANNHEIM GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .60
                     /organism="unidentified"
                     /db_xref="taxon:32644"
```

```
BASE COUNT        0 a     30 c      0 g      0 t     30 others
ORIGIN

  Query Match           10.1%;  Score 24;  DB 25;  Length 60;
  Best Local Similarity   6.2%;  Pred. No. 1.07e+00;
  Matches     2;  Conservative   26;  Mismatches   4;  Indels   0;  Gaps   0;

Db      29 CCRRRRRRRRRRRRRRRRRRRRRRRRRRRR 60
           ||:::::::  :::::::::: :::::: ::
Cp      94 CCAAGGAGGGTCAAGAGGAGGACAAGGGGCAG 63


RESULT   14
LOCUS       MVU87256     1056 bp    DNA             MAM       02-JAN-1999
DEFINITION  Mustela vison GT dinucleotide repeat, chromosome 1q.
ACCESSION   U87256
NID         g4099442
VERSION     U87256.1  GI:4099442
KEYWORDS    .
SOURCE      American mink.
  ORGANISM  Mustela vison
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Carnivora; Fissipedia; Mustelidae; Mustela.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Brusgaard,K., Shukri,N.M., Malchenko,S., Koroleva,I. and Lohi,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of
            Animal Science, Blichersalle K25, Tjele 8830, DK
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mustela vison"
                     /db_xref="taxon:9667"
                     /chromosome="1"
                     /map="1q"
                     /note="primers: 1167F: agcccctgcatatctacttctt, 1167R:
                     gaggatcttaccgctgttgag"
     primer_bind     98. .119
                     /standard_name="1167F"
     primer_bind     complement(300. .320)
                     /standard_name="1167R"
BASE COUNT      211 a    221 c    210 g    225 t    189 others
ORIGIN

  Query Match           10.1%;  Score 24;  DB 23;  Length 1056;
  Best Local Similarity  10.8%;  Pred. No. 1.07e+00;
  Matches     7;  Conservative   40;  Mismatches  17;  Indels   1;  Gaps   1;

Db     611 KRMRRCSRYMMMRMCVGSGTWARCCCCDKKSGSGDKSHCKSKRKYKWMDRYHBC-KSMCA 669
           ::::: : :: : :|:|:  :   || :::: : ::::  :::: ::::: ::  :::||
Cp     130 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 71

Db     670 MVRGK 674
           :::|:
Cp      70 AGGGG 66


RESULT   15
LOCUS       AC005920   175793 bp    DNA             PRI       25-NOV-1998
DEFINITION  Homo sapiens chromosome 17, clone hRPK.700_H_6, complete sequence.
ACCESSION   AC005920
NID         g3927856
VERSION     AC005920.1  GI:3927856
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175793)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.700_H_6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175793)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 175793)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 25, 1998 this sequence version replaced gi:3924659.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .175793
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.700_H_6"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   complement(1. .80)
                     /rpt_family="AluJ/FLAM"
     repeat_region   1575. .1689
                     /rpt_family="L2"
     repeat_region   1745. .1868
                     /rpt_family="L2"
     repeat_region   2056. .2266
                     /rpt_family="MER3"
     repeat_region   2552. .2868
                     /rpt_family="AluJo"
     repeat_region   3522. .3817
                     /rpt_family="AluSx"
     repeat_region   4183. .4775
                     /rpt_family="L1ME1"
     repeat_region   complement(5087. .5148)
                     /rpt_family="7SK"
     repeat_region   6025. .6165
                     /rpt_family="MIR"
     repeat_region   complement(6319. .6602)
                     /rpt_family="AluSg"
     repeat_region   7218. .7519
                     /rpt_family="AluSx"
     repeat_region   7520. .7852
                     /rpt_family="AluSc"
     repeat_region   7853. .7936
```

```
                     /rpt_family="L2"
     repeat_region   complement(8343. .8418)
                     /rpt_family="MIR"
     repeat_region   8478. .8847
                     /rpt_family="L2"
     repeat_region   complement(8957. .9087)
                     /rpt_family="MIR"
     repeat_region   complement(9088. .9389)
                     /rpt_family="AluSx"
     repeat_region   complement(10229. .10415)
                     /rpt_family="MIR"
     repeat_region   10578. .10798
                     /rpt_family="MIR"
     repeat_region   11365. .11656
                     /rpt_family="AluSg1"
     repeat_region   complement(12745. .12793)
                     /rpt_family="MER33"
     repeat_region   complement(12794. .13103)
                     /rpt_family="AluSx"
     repeat_region   complement(13104. .13377)
                     /rpt_family="MER33"
     repeat_region   13536. .13643
                     /rpt_family="MIR"
     repeat_region   complement(13867. .13909)
                     /rpt_family="A-rich"
     repeat_region   complement(13912. .14209)
                     /rpt_family="AluSx"
     repeat_region   complement(14573. .14604)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(14606. .14884)
                     /rpt_family="AluSg"
     repeat_region   complement(15772. .15816)
                     /rpt_family="(GGGGA)n"
     repeat_region   complement(15817. .16101)
                     /rpt_family="AluSg"
     repeat_region   complement(16102. .16115)
                     /rpt_family="(GGGGA)n"
     repeat_region   complement(16116. .16359)
                     /rpt_family="AluJo"
     repeat_region   complement(16533. .16579)
                     /rpt_family="MIR"
     repeat_region   complement(16855. .16946)
                     /rpt_family="MIR"
     repeat_region   17505. .17532
                     /rpt_family="GC_rich"
     repeat_region   complement(19845. .19992)
                     /rpt_family="MIR"
     repeat_region   complement(20570. .20601)
                     /rpt_family="L2"
     repeat_region   complement(20617. .20643)
                     /rpt_family="(GAAA)n"
     repeat_region   complement(20648. .20927)
                     /rpt_family="AluSx"
     repeat_region   complement(21024. .21122)
                     /rpt_family="MIR"
     repeat_region   complement(21690. .21999)
                     /rpt_family="AluY"
     repeat_region   complement(22181. .22464)
                     /rpt_family="AluJb"
     repeat_region   complement(24269. .24408)
                     /rpt_family="MIR"
     repeat_region   complement(24707. .25000)
                     /rpt_family="AluJo"
     repeat_region   complement(25393. .25453)
                     /rpt_family="L2"
     repeat_region   complement(25736. .25870)
                     /rpt_family="MER5A"
     repeat_region   complement(25867. .26026)
                     /rpt_family="MER5B"
     repeat_region   26623. .26707
                     /rpt_family="MIR"
     repeat_region   26992. .27200
                     /rpt_family="MIR"
     repeat_region   complement(27291. .27579)
                     /rpt_family="AluSx"
     repeat_region   27822. .28002
                     /rpt_family="MIR"
     repeat_region   27886. .28014
                     /rpt_family="L2"
     repeat_region   28023. .28090
                     /rpt_family="MIR"
     repeat_region   28051. .28103
                     /rpt_family="L2"
     repeat_region   complement(28331. .28368)
                     /rpt_family="(TGAA)n"
     repeat_region   28956. .29258
                     /rpt_family="L2"
     repeat_region   complement(29275. .29774)
                     /rpt_family="MLT1G"
     repeat_region   complement(29874. .30088)
                     /rpt_family="AluSg/x"
     repeat_region   complement(30799. .31109)
                     /rpt_family="AluSx"
     repeat_region   32152. .32205
                     /rpt_family="AluJ/FRAM"
     repeat_region   32214. .32270
                     /rpt_family="(GA)n"
     repeat_region   complement(32584. .33089)
                     /rpt_family="L1M3e"
     repeat_region   33605. .33898
                     /rpt_family="AluSg"
     repeat_region   complement(33919. .34193)
                     /rpt_family="MER7A"
     repeat_region   34194. .34480
                     /rpt_family="AluYb8"
     repeat_region   34481. .34501
                     /rpt_family="(TAA)n"
     repeat_region   complement(34502. .34565)
                     /rpt_family="MER7A"
     repeat_region   complement(36171. .36324)
                     /rpt_family="L2"
...
Note: remainder of annotations omitted.

  Query Match           10.1%;  Score 24;  DB 31;  Length 175793;
  Best Local Similarity  72.3%;  Pred. No. 1.07e+00;
  Matches    47;  Conservative    0;  Mismatches  17;  Indels   1;  Gaps   1;

Db  18839 GGGAGCCAAGAGCCCAGAGGGGTCACTTGGGCCCAGGGAGGGTGG-GACACAGACGTCTG 18897
          || |||| ||| |||| || || |||| | || |||| || || ||||| || |||
Cp     67 GGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATGGAGGAGACACTGAAGTCCC 8

Db  18898 GGTGG 18902
          |||||
Cp      7 GGTGG 3


Search completed: Sun Oct 24 16:30:24 1999
Job time : 485 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 16:38:53 1999; MasPar time 73.89 Seconds
693.101 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-1
Description: (1-239) from US09092296.seq
Perfect Score: 238
N.A. Sequence: 1 GGCCACCGGGACTTCAGTGT.....CCCTTCAGGGACCAGCGTCA 239
Comp: CCGGTGGCCCTGAAGTCACA.....GGGAAGTCCCTGGTCGCAGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 7.797; Variance 4.949; scale 1.575

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	232	97.5	439	60	V84366	Human stomach carcino	1.08e-132
c	2	37	15.5	91	9	Q51746	Oligonucleotide probe	3.39e-08
	3	37	15.5	204	1	N81164	Base substituted E.co	3.39e-08
c	4	37	15.5	204	1	N81164	Base substituted E.co	3.39e-08
	5	36	15.1	91	9	Q51746	Oligonucleotide probe	1.21e-07
	6	31	13.0	114	12	Q70470	Generic DNA sequence	6.14e-05
c	7	31	13.0	114	12	Q70467	Generic DNA sequence	6.14e-05
	8	30	12.6	114	12	Q70468	Generic DNA sequence	2.07e-04
c	9	30	12.6	114	12	Q70470	Generic DNA sequence	2.07e-04
c	10	30	12.6	114	12	Q70468	Generic DNA sequence	2.07e-04
	11	29	12.2	114	12	Q70467	Generic DNA sequence	6.89e-04
	12	29	12.2	114	12	Q70465	Generic DNA sequence	6.89e-04
	13	29	12.2	114	12	Q70469	Generic DNA sequence	6.89e-04
c	14	28	11.8	114	12	Q70465	Generic DNA sequence	2.27e-03
	15	28	11.8	178	32	T76405	Human endothelin-1 an	2.27e-03
	16	27	11.3	114	12	Q70466	Generic DNA sequence	7.35e-03
	17	27	11.3	172	32	T76363	Human interleukin 8 a	7.35e-03
	18	26	10.9	114	12	Q70472	Generic DNA sequence	2.35e-02
c	19	26	10.9	114	12	Q70469	Generic DNA sequence	2.35e-02
c	20	25	10.5	39	7	Q51787	Mixed oligonucleotide	7.40e-02
c	21	25	10.5	114	12	Q70472	Generic DNA sequence	7.40e-02
	22	25	10.5	250	32	T76438	Substance P antisense	7.40e-02
	23	24	10.1	114	12	Q70473	Generic DNA sequence	2.29e-01
c	24	24	10.1	114	12	Q70466	Generic DNA sequence	2.29e-01
c	25	24	10.1	114	12	Q70473	Generic DNA sequence	2.29e-01
	26	23	9.7	114	12	Q70471	Generic DNA sequence	6.97e-01
c	27	23	9.7	114	12	Q70471	Generic DNA sequence	6.97e-01
	28	23	9.7	565	6	Q35072	HCV envelope region n	6.97e-01
c	29	22	9.2	36	2	Q11195	Ballast Constituent c	2.08e+00
	30	22	9.2	75	21	T13612	DC43 TSAR library gen	2.08e+00
c	31	22	9.2	81	21	T13611	DC43 TSAR library gen	2.08e+00
	32	22	9.2	82	21	T13610	DC43 TSAR library gen	2.08e+00
	33	22	9.2	89	32	T76219	Human IL5 antisense o	2.08e+00
c	34	22	9.2	91	46	V44650	Mammalian DNA replica	2.08e+00
	35	22	9.2	190	32	T76452	Chymase antisense oli	2.08e+00
	36	22	9.2	264	32	T76445	Substance P receptor	2.08e+00
c	37	22	9.2	3088	16	T05628	ADP ribosylation fact	2.08e+00
	38	21	8.8	168	32	T76270	Human MDNCF antisense	6.06e+00
	39	21	8.8	908	51	V63025	B. malayi ankyrin cDN	6.06e+00
c	40	21	8.8	908	52	V63315	Nucleotide nBmAnk908	6.06e+00
c	41	21	8.8	908	51	V63024	B. malayi ankyrin nBm	6.06e+00
c	42	21	8.8	909	51	V63014	D. immitis ankyrin nD	6.06e+00
c	43	21	8.8	911	52	V63312	Nucleotide nDiAnk911	6.06e+00
c	44	21	8.8	911	51	V63012	D. immitis ankyrin nD	6.06e+00
c	45	21	8.8	5503	51	V63020	D. immitis ankyrin nD	6.06e+00

ALIGNMENTS

RESULT 1
ID V84366 standard; cDNA to mRNA; 439 BP.
AC V84366;
DT 30-MAR-1999 (first entry)
DE Human stomach carcinoma cDNA clone HP10408.
KW Transmembrane protein; HP10408; human; stomach cancer; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 75..311
FT /*tag= a
FT /note= "cDNA comprising the coding region (minus
FT the stop codon) is claimed (Claim 3)"
PN WO9855508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI; 99-045730/04.
DR P-PSDB; W88498.
PT New human proteins containing transmembrane domains and their
PT encoding sequences - useful in the preparation of antibodies and
PT large-scale protein production, gene diagnosis, and gene therapy
PS Claim 4; Page 135; 178pp; English.
CC This is the nucleotide sequence of cDNA clone HP10408, which
CC includes a coding region (also claimed) for a novel human
CC transmembrane protein (see W88498). The clone was isolated from a
CC stomach cancer cDNA library using a signal sequence detection
CC method, and by protein synthesis by in vitro translation. The
CC encoded protein has a putative signal sequence and a putative
CC internal transmembrane domain. The invention provides nucleotide
CC sequences (see V84359-76) coding for 18 transmembrane proteins

```
CC   (see W88491-508), vectors containing such polynucleotides, and
CC   eukaryotic cells containing the vectors.  The proteins can be
CC   used as antigens or as compositions in the preparation of
CC   antibodies against the proteins.  The polynucleotides can be used
CC   as probes for gene diagnosis, and as gene sources for gene therapy
CC   and large-scale production of proteins encoded by the cDNA.  The
CC   host cells are used for the detection of ligands corresponding to
CC   the expressed proteins, and the screening of low mol.wt. medicines.
SQ   Sequence   439 BP;    89 A;    137 C;    109 G;    104 T;

  Query Match           97.5%;  Score 232;  DB 60;  Length 439;
  Best Local Similarity  99.2%;  Pred. No. 1.08e-132;
  Matches   238;  Conservative    0;  Mismatches   1;  Indels   1;  Gaps   1;

Db     25 ggccaccgggacttcagtgtctcctccatcccaggagcgcagtggccactatggggtctg 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 GGCCACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTG 60

Db     85 ggctgccccttgtcctcctcttgaccctccttggcagctcacatggaacagggccgggta 144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 GGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTA 120

Db    145 tgactttgcaactgaagctgaaggagtctttttctgacaaattcctcctatgagtccagct 204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 TGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCT 180

Db    205 tcctggaattgcttgaaaagctctgcctcctcctccatctcccttcagggaccagcgtca 264
          ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Qy    181 TCCTGGAATTGCTTGAAAAN-TCTGCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCA 239


RESULT    2
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match           15.5%;  Score 37;  DB 9;  Length 91;
  Best Local Similarity   5.7%;  Pred. No. 3.39e-08;
  Matches     3;  Conservative   42;  Mismatches   8;  Indels   0;  Gaps   0;

Db     11 ssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsvctc 63
          :::::  :: ::: :: :  :::::::::::::::::::: :::::::::|||
Cp     87 GGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTC 35


RESULT    3
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match           15.5%;  Score 37;  DB 1;  Length 204;
  Best Local Similarity   8.7%;  Pred. No. 3.39e-08;
  Matches    10;  Conservative   59;  Mismatches  46;  Indels   0;  Gaps   0;

Db     72 gaayycdchvgccgymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynach 131
          || ::|: :: |  :    |::::: :: ::::: :: |  ::   : : : :|    |:
Qy     10 GACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCC 69

Db    132 hddhyvybbbvynvhnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
          :::::::::::: :: :  |    :   : : :::  :: :  : :: : ::::|
Qy     70 TTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 124


RESULT    4
ID   N81164 standard; DNA; 204 BP.
AC   N81164;
DT   08-NOV-1990  (first entry)
DE   Base substituted E.coli beta-galactosidase alpha-fragment.
KW   E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS   Escherichia coli.
FH   Key             Location/Qualifiers
FT   misc_feature    19..69
FT                   /*tag=  a
FT                   /function=multiple cloning site
FT   primer_bind     187..204
FT                   /*tag=  b
PN   EP-285123-A.
PD   05-MAY-1988.
PF   30-MAR-1988; 105163.
PR   03-APR-1987; US-034819.
PA   (SUSO) SUOMEN SOKERI OY.
PI   Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR   WPI; 88-279927/40.
PT   Introducing random point mutations into nucleic acods -
PT   by prepn of single stranded template, annealing a primer, elongation,
PT   misincorporation, completion of molecules and screening.
PS   Disclosure; p; English.
CC   Random point mutations were introduced into the alpha fragment of
CC   E.coli beta-galactosidase. The wild type sequence was obtained as a
CC   single stranded template and an oligonucleotide was hybridised to
CC   it to generate a popn of DNA molecules which terminate at all
CC   possible nucleotide positions within a specified region. The
```

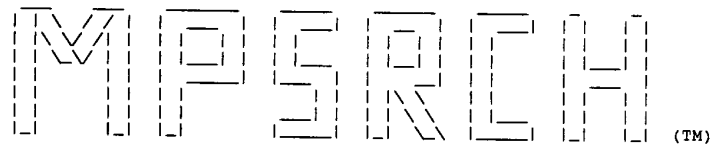
```
CC   variable 3' ends generated in this way are used as primers for
CC   reverse transcriptase. Nucleotides are misincorporated by the
CC   transcriptase and the molecules are completed to forms that can be
CC   amplified and then expressed in a suitable host-vector system.
CC   The sequence covers all 176 difft base substitutions, most of which
CC   occurred singularly in any given mutant.
CC   See also P80575.
SQ   Sequence   204 BP;   21 A;   47 C;   17 G;   11 T;   108 Others;

  Query Match           15.5%;  Score 37;  DB 1;  Length 204;
  Best Local Similarity   7.9%;  Pred. No. 3.39e-08;
  Matches     8;  Conservative   54;  Mismatches  39;  Indels   0;  Gaps   0;

Db     86 ymrttthhyrrmrbnvyrdynrsdaaawyccyrrsvkydccynachhddhyvybbbvynv 145
          :::|   : ::: : :: :  : : | ::| : :::::   :  |:::::::: :::  :
Cp    172 TCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGC 113

Db    146 hnhnncncccbnnhvchnvhbnnhrnwayvrhdarrddvhc 186
          : :    |  :  ::|: ::    :   :::::|:::::::|
Cp    112 CCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGAC 72


RESULT    5
ID   Q51746 standard; cDNA; 91 BP.
AC   Q51746;
DT   31-MAY-1994  (first entry)
DE   Oligonucleotide probe MK14-A
KW   Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW   ss.
OS   Synthetic.
PN   EP-571911-A.
PD   01-DEC-1993.
PF   24-MAY-1993; 108325.
PR   26-MAY-1992; US-889651.
PA   (BECT ) BECTON DICKINSON CO.
PI   Shank DD,  Spears PA;
DR   WPI; 93-378844/48.
PT   New oligo:nucleotide probes specific for Mycobacteria - used for
PT   detection and amplification of Mycobacteria nucleic acid in
PT   samples
PS   Claim 3; Page 14; 23pp; English.
CC   Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC   (Q51735).  It hybridized to all spp. of mycobacteria tested, but
CC   cross reacted to a few non-mycobacterial spp.  The probe may
CC   be useful as an initial screen for mycobacterial infection.
CC   See also Q51735-45 and Q51747-59.
SQ   Sequence   91 BP;    5 A;    17 C;    15 G;    4 T;

  Query Match           15.1%;  Score 36;  DB 9;  Length 91;
  Best Local Similarity   0.0%;  Pred. No. 1.21e-07;
  Matches     0;  Conservative   43;  Mismatches   7;  Indels   0;  Gaps   0;

Db     11 ssvhsyyvvhvvshhhsvhhvvhhvhvsvvvvhhvvhvvhhvhyhvyvsv 60
          :::  :: ::: :::::::::::::::::::: ::::::::::::  :::
Qy      7 CGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGG 56


RESULT    6
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    5 A;    10 C;    0 G;    0 T;

  Query Match           13.0%;  Score 31;  DB 12;  Length 114;
  Best Local Similarity   6.8%;  Pred. No. 6.14e-05;
  Matches     7;  Conservative   28;  Mismatches  68;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbcacnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :     :||   :  :  :  :|||  :  :  :  :  :  :  :  :
Qy     22 TCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCT 81

Db     63 bnnbnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnb 105
          :  :  :  :  :  :| |     :  :  :  :  :  :  :
Qy     82 TGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 124


RESULT    7
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
```

```
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           13.0%;  Score 31;  DB 12;  Length 114;
  Best Local Similarity   4.6%;  Pred. No. 6.14e-05;
  Matches     5;  Conservative   30;  Mismatches  74;  Indels   0;  Gaps   0;

Db      6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnnbnn 65
          :  :  :  :  :  :  :     :  :  :  :  :  :  :  :||   :       :
Cp    144 CCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGG 85

Db     66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnb 114
          :     :  :  :  :  :  :  :     :  :  :  :  :|||  :
Cp     84 TCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCT 36


RESULT    8
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           12.6%;  Score 30;  DB 12;  Length 114;
  Best Local Similarity   3.8%;  Pred. No. 2.07e-04;
  Matches     4;  Conservative   30;  Mismatches  72;  Indels   0;  Gaps   0;

Db      9 bnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnnbnnbnn 68
          :  :  :     :  :  :  :  :||  :  :  :  :  :  :        :  :
Qy     19 GTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCC 78

Db     69 bnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
          :  :  :  :  :||   :  :     :  :  :  :  :  :  :
Qy     79 TCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 124


RESULT    9
ID   Q70470 standard; DNA; 114 BP.
AC   Q70470;
DT   10-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concatenated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "encoded by Z (see comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R58378.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 36; 255pp; English.
CC   Q70470 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8
CC   -(CAC)2(NNB)Y. X and Y are flanking  restriction sites (X is not the same
CC   as Y) that are not specified further. The peptides generated by this and
CC   other generic sequences (Q70471-73) have invariant histidine residues
CC   incorporated into variant sequences. TSARs are concatenated
CC   heterofunctional proteins or peptides, comprising at least two functional
CC   regions - a binding domain with affinity for a ligand and a second
CC   effector peptide portion that is chemically or biologically active.They
CC   may further comprise a linker peptide between the 2 domains. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
```

```
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;     5 A;     10 C;     0 G;     0 T;

  Query Match           12.6%;  Score 30;  DB 12;  Length 114;
  Best Local Similarity   6.7%;  Pred. No. 2.07e-04;
  Matches     7;  Conservative   28;  Mismatches  69;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbcacnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           :  :  :  :||   :  :  :  :| |  :  :  :     :  :  :  :
Cp     138 GCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGA 79

Db      63 bnnbnnbnnbnnbnnbcacnnbnnbnnbnnbnnbnnbnnbnnbc 106
           :  :  :  :  :  :| |     :  :  :  :  :  :  :|
Cp      78 GGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTC 35


RESULT   10
ID   Q70468 standard; DNA; 114 BP.
AC   Q70468;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65154.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70468 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed activity
CC   allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;     0 A;     2 C;     2 G;     2 T;

  Query Match           12.6%;  Score 30;  DB 12;  Length 114;
  Best Local Similarity   2.7%;  Pred. No. 2.07e-04;
  Matches     3;  Conservative   32;  Mismatches  77;  Indels   0;  Gaps   0;

Db       3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           :  :  :  :     :  :  :  :  :  :  :| |  :     :  :  :  :
Cp     138 GCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGA 79

Db      63 bnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
           :  :  :  :  :  :  :  :| |  :  :  :  :  :  :  :  :  :
Cp      78 GGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATG 27


RESULT   11
ID   Q70467 standard; DNA; 114 BP.
AC   Q70467;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65153.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70467 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)16(TGC)(NNB)1Z(NNB)16(TGC)(NNB)1Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need for
CC   complex methods of hybridoma formation or in vivo antibody production.
CC   The TSARs are easily characterised and have designed activity allowing
CC   direct and rapid detection in a screening process.
SQ   Sequence   114 BP;     0 A;     2 C;     2 G;     2 T;

  Query Match           12.2%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   3.8%;  Pred. No. 6.89e-04;
  Matches     4;  Conservative   29;  Mismatches  72;  Indels   0;  Gaps   0;

Db       6 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnnnnnnnbnn 65
           :  :  :  :  :     :  :  :     :  :  :  :  :  :|| :        :
Qy       1 GGCCACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTG 60
```

```
Db      66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtg 110
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  ||
Qy      61 GGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATG 105


RESULT   12
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;  0 A;    2 C;    2 G;    2 T;

  Query Match           12.2%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   2.1%;  Pred. No. 6.89e-04;
  Matches     2;  Conservative   29;  Mismatches  66;  Indels   0;  Gaps   0;

Db       6 bnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnnnbnn 65
           :  :  :  :  :  ||  :  :  :  :  :  :  :  :  :  :  :  :  :
Qy      31 CCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCC 90

Db      66 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 102
           :  :  :  :     :  :  :  :  :  :  :  :
Qy      91 TTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTT 127


RESULT   13
ID   Q70469 standard; DNA; 114 BP.
AC   Q70469;
DT   07-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR peptide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag=  a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6,9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70469 is a generic DNA sequence used to generate random TSAR peptide
CC   This generic formula can be represented  as follows: X(TGC)(NNB)10-
CC   (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking  restriction
CC   sites (X is not the same as Y) that are not specified further. This
CC   sequence generates peptides that are cloverleaf in structure. Other
CC   generic sequences are shown in Q70465-68. Other specific peptides
CC   generated by these generic sequences are shown in R65150-54. TSARs are
CC   concatenated heterofunctional proteins or peptides, comprising at least
CC   two functional regions - a binding domain with affinity for a ligand and
CC   a second effector peptide portion that is chemically or biologically
CC   active.They may further comprise a linker peptide between the 2 domains.
CC   The oligonucleotides are also designed so that the expressed peptide
CC   contains 2 or 4 cysteine residues positioned in, or flanking, the
CC   unpredicted or variant residues. These residues confer some degree of
CC   conformational rigidity to the peptides. The TSARs or compsns. comprising
CC   a TSAR binding domain can be used in vivo to deliver a chemically or
CC   biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
CC   or enzyme, to the specific target or on the cell. They can also replace
CC   the function of macromolecules, eg. monoclonal or polyclonal antibodies
CC   and therefore circumvent the need for complex methods of hybridoma
CC   formation or in vivo antibody production. The TSARs are easily
CC   characterised and have designed activity allowing direct and rapid
CC   detection in a screening process.
SQ   Sequence   114 BP;    0 A;    4 C;    4 G;    4 T;

  Query Match           12.2%;  Score 29;  DB 12;  Length 114;
  Best Local Similarity   4.6%;  Pred. No. 6.89e-04;
  Matches     5;  Conservative   29;  Mismatches  75;  Indels   0;  Gaps   0;

Db       3 cnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnnnnnnn 62
           |  :  :     :  :  :     :  :  :  ||  :  :  :  :  :  :  :
Qy       7 CGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGC 66

Db      63 bnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 111
           :  :||   :  :  :  :  :  :  :  :  :     :  :  :  :
Qy      67 CCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCC 115


RESULT   14
ID   Q70465 standard; DNA; 114 BP.
AC   Q70465;
DT   05-APR-1995  (first entry)
DE   Generic DNA sequence to generate a random TSAR petide library.
KW   TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW   effector domain; concateneated heterofunctional protein; linker;
KW   direct; rapid; detection; screening; treatment; generic; ss.
OS   Synthetic.
```

```
FH   Key             Location/Qualifiers
FT   misc_feature    55..60
FT                   /*tag= a
FT                   /note= "this sequence represents 'Z'; Z can be a
FT                   sequence of 6, 9 or 12 nucleotides (see
FT                   comments)"
PN   WO9418318-A.
PD   18-AUG-1994.
PF   01-FEB-1994; U00977.
PR   01-FEB-1993; US-013416.
PR   30-DEC-1993; US-176500.
PR   31-JAN-1994; US-189331.
PA   (UYNC-) UNIV NORTH CAROLINA.
PI   Fowlkes DM,  Kay BK;
DR   WPI; 94-279739/34.
DR   P-PSDB; R65150 and R65151.
PT   Identifying proteins or peptide(s) which bind a ligand - by
PT   screening a recombinant vector library expressing fusion proteins
PT   comprising a binding domain and an effector domain
PS   Disclosure; Page 35; 255pp; English.
CC   Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC   Synthetic Affinity Reagents) peptides.This generic formula can also be
CC   represented  as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X
CC   and Y are flanking  restriction sites (X is not the same as Y) that are
CC   not specified further. Other generic sequences are shown in Q70466-68.
CC   Other specific peptides generated by these generic sequences are shown in
CC   R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC   comprising at least two functional regions - a binding domain with
CC   affinity for a ligand and a second effector peptide portion that is
CC   chemically or biologically active.They may further comprise a linker
CC   peptide between the 2 domains.The oligonucleotides are also designed so
CC   that the expressed peptide contains 2 or 4 cysteine residues positioned
CC   in, or flanking, the unpredicted or variant residues. These residues
CC   confer some degree of conformational rigidity to the peptides. The TSARs
CC   or compsns. comprising a TSAR binding domain can be used in vivo to
CC   deliver a chemically or biologically active moiety, eg. metal ion,
CC   radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC   cell. They can also replace the function of macromolecules, eg.
CC   monoclonal or polyclonal antibodies and therefore circumvent the need
CC   for complex methods of hybridoma formation or in vivo antibody
CC   production. The TSARs are easily characterised and have designed
CC   activity allowing direct and rapid detection in a screening process.
SQ   Sequence   114 BP;    0 A;    2 C;    2 G;    2 T;

  Query Match           11.8%;  Score 28;  DB 12;  Length 114;
  Best Local Similarity  3.6%;  Pred. No. 2.27e-03;
  Matches     4;  Conservative   30;  Mismatches  78;  Indels   0;  Gaps   0;

Db      3 bnnbnnbnnbnnbnnbtgcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnn 62
          :  :  :  :  :    |  :  :  :  :  :  :  :     :  :  :
Cp    141 TCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCA 82

Db     63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbtgcnnbnnbnnb 114
           :  :  :  :  :  :  :     :  :  :  :  :|||  :  :  :
Cp     81 AGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGG 30


RESULT   15
ID   T76405 standard; DNA; 178 BP.
AC   T76405;
DT   15-SEP-1997  (first entry)
DE   Human endothelin-1 antisense oligonucleotide.
KW   Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW   chronic obstructive pulmonary disease; bronchitis; ss.
OS   Synthetic.
PN   WO9640162-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09306.
PR   07-JUN-1995; US-474497.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Metzger WJ,  Nyce JW;
DR   WPI; 97-051871/05.
PT   Treatment of airway diseases such as asthma - by topically applying
PT   adenosine-free antisense oligo:nucleotide to airway epithelium of
PT   subject
PS   Claim 5; Page 38; 71pp; English.
CC   A method for treating airway disease in a subject has been produced,
CC   which involves the topical administration of an essentially adenosine
CC   free antisense oligonucleotide (ON) to the airway epithelium of the
CC   subject. The present sequence is an antisense oligonucleotide specific
CC   for the human endothelin-1, targeted at the initiation codon. The
CC   method can be used to treat airway diseases such as cystic fibrosis,
CC   asthma, chronic obstructive pulmonary disease, bronchitis and other
CC   airway diseases characterised by an inflammatory response. By
CC   eliminating adenosine from the antisense ON, its liberation upon
CC   antisense degradation is prevented, thereby preventing adenosine-
CC   induced bronchoconstriction in patients with hyper-reactive airways.
SQ   Sequence   178 BP;    0 A;    52 C;    46 G;    32 T;

  Query Match           11.8%;  Score 28;  DB 32;  Length 178;
  Best Local Similarity  32.0%;  Pred. No. 2.27e-03;
  Matches    32;  Conservative   37;  Mismatches  29;  Indels   2;  Gaps   2;

Db     71 bgcgcctbbgbctgctgtttctggbgc-tccttggcbbgccbcbbbc-bgcbgbgbgbbb 128
            :||  | ::| || | || |||  : | |||  || ::|| : :::| :  :|:|: :::
Qy      1 GGCCACCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTG 60

Db    129 btcbtgbgcbbbtbbtccbttctgbbbbbbbgggbtcbbb 168
            : |:  : |:::|::|||: |   ::::::: ||:  :::
Qy     61 GGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTC 100


Search completed: Sun Oct 24 16:40:11 1999
Job time : 78 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 16:40:29 1999; MasPar time 21.17 Seconds
976.635 Million cell updates/sec
Tabular output not generated.

Title: >US-09-092-296-1
Description: (1-239) from US09092296.seq
Perfect Score: 238
N.A. Sequence: 1 GGCCACCGGGACTTCAGTGT.....CCCTTCAGGGACCAGCGTCA 239
Comp: CCGGTGGCCCTGAAGTCACA.....GGGAAGTCCCTGGTCGCAGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 165359 seqs, 43243793 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-issued
1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 7.391; Variance 4.188; scale 1.765

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description	Pred. No.
1		47	19.7	7218	2	US-08-232-	Sequence 14, Applicati	8.84e-17
c 2		36	15.1	215	1	US-08-238-	Sequence 5, Applicatio	1.01e-09
3		31	13.0	965	3	US-08-388-	Sequence 22, Applicati	1.18e-06
4		27	11.3	215	1	US-08-238-	Sequence 5, Applicatio	2.69e-04
c 5		26	10.9	965	3	US-08-388-	Sequence 22, Applicati	1.01e-03
6		22	9.2	75	4	PCT-US95-1	Sequence 99, Applicati	1.67e-01
c 7		22	9.2	81	4	PCT-US95-1	Sequence 98, Applicati	1.67e-01
8		22	9.2	82	4	PCT-US95-1	Sequence 97, Applicati	1.67e-01
c 9		22	9.2	3088	3	US-08-418-	Sequence 1, Applicatio	1.67e-01
c 10		21	8.8	65	1	US-08-471-	Sequence 145, Applicat	5.69e-01
11		21	8.8	66	1	US-08-471-	Sequence 144, Applicat	5.69e-01
c 12		21	8.8	68	1	US-08-471-	Sequence 143, Applicat	5.69e-01
13		21	8.8	69	1	US-08-471-	Sequence 142, Applicat	5.69e-01
14		21	8.8	74	4	PCT-US95-1	Sequence 94, Applicati	5.69e-01
c 15		21	8.8	74	4	PCT-US95-1	Sequence 100, Applicat	5.69e-01
c 16		21	8.8	75	4	PCT-US95-1	Sequence 99, Applicati	5.69e-01
17		21	8.8	81	4	PCT-US95-1	Sequence 98, Applicati	5.69e-01
18		21	8.8	81	4	PCT-US95-1	Sequence 92, Applicati	5.69e-01
c 19		21	8.8	82	4	PCT-US95-1	Sequence 97, Applicati	5.69e-01
20		21	8.8	906	3	US-09-031-	Sequence 41, Applicati	5.69e-01
	21	21	8.8	906	3	US-08-847-	Sequence 41, Applicati	5.69e-01
c	22	21	8.8	906	3	US-08-847-	Sequence 40, Applicati	5.69e-01
c	23	21	8.8	906	3	US-09-031-	Sequence 40, Applicati	5.69e-01
	24	21	8.8	908	3	US-09-031-	Sequence 39, Applicati	5.69e-01
	25	21	8.8	908	3	US-08-847-	Sequence 39, Applicati	5.69e-01
c	26	21	8.8	908	3	US-08-847-	Sequence 37, Applicati	5.69e-01
c	27	21	8.8	908	3	US-09-031-	Sequence 37, Applicati	5.69e-01
	28	21	8.8	909	3	US-08-847-	Sequence 26, Applicati	5.69e-01
	29	21	8.8	909	3	US-09-031-	Sequence 26, Applicati	5.69e-01
c	30	21	8.8	909	3	US-09-031-	Sequence 25, Applicati	5.69e-01
c	31	21	8.8	909	3	US-08-847-	Sequence 25, Applicati	5.69e-01
	32	21	8.8	911	3	US-09-031-	Sequence 24, Applicati	5.69e-01
	33	21	8.8	911	3	US-08-847-	Sequence 24, Applicati	5.69e-01
c	34	21	8.8	911	3	US-08-847-	Sequence 22, Applicati	5.69e-01
c	35	21	8.8	911	3	US-09-031-	Sequence 22, Applicati	5.69e-01
	36	21	8.8	5235	3	US-09-031-	Sequence 36, Applicati	5.69e-01
	37	21	8.8	5235	3	US-08-847-	Sequence 36, Applicati	5.69e-01
c	38	21	8.8	5235	3	US-09-031-	Sequence 35, Applicati	5.69e-01
c	39	21	8.8	5235	3	US-08-847-	Sequence 35, Applicati	5.69e-01
	40	21	8.8	5503	3	US-09-031-	Sequence 34, Applicati	5.69e-01
	41	21	8.8	5503	3	US-08-847-	Sequence 34, Applicati	5.69e-01
c	42	21	8.8	5503	3	US-08-847-	Sequence 32, Applicati	5.69e-01
c	43	21	8.8	5503	3	US-09-031-	Sequence 32, Applicati	5.69e-01
c	44	20	8.4	66	4	PCT-US95-1	Sequence 93, Applicati	1.89e+00
c	45	20	8.4	1185	1	US-08-049-	Sequence 5, Applicatio	1.89e+00

ALIGNMENTS

```
RESULT    1
ID   US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC   xxxxxx
DT
DE   Sequence 14, Application US/08232463
CC   Sequence 14, Application US/08232463
CC   Patent No. 5670367
CC    GENERAL INFORMATION:
CC      APPLICANT:  DORNER, F.
CC      APPLICANT:  SCHEIFLINGER, F.
CC      APPLICANT:  FALKNER, F. G.
CC      TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
CC      NUMBER OF SEQUENCES:  52
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Foley & Lardner
CC        STREET:  1800 Diagonal Road, Suite 500
CC        CITY:  Alexandria
CC        STATE:  VA
CC        COUNTRY:  USA
CC        ZIP:  22313-0299
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/232,463
CC        FILING DATE:
CC        CLASSIFICATION:  435
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  US/07/935,313
CC        FILING DATE:
CC        APPLICATION NUMBER:  EP 91 114 300.6
CC        FILING DATE:  26-AUG-1991
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  BENT, Stephen A.
CC        REGISTRATION NUMBER:  29,768
CC        REFERENCE/DOCKET NUMBER:  30472/114 IMMU
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703)836-9300
CC        TELEFAX:  (703)683-4109
CC        TELEX:  899149
CC    INFORMATION FOR SEQ ID NO:  14:
CC      SEQUENCE CHARACTERISTICS:
```

```
CC       LENGTH:  7218 base pairs
CC       TYPE:  nucleic acid
CC       STRANDEDNESS:  single
CC       TOPOLOGY:  linear
CC     IMMEDIATE SOURCE:
CC       CLONE:  pTZgpt-Fls
SQ   SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

  Query Match           19.7%;  Score 47;  DB 2;  Length 7218;
  Best Local Similarity   2.3%;  Pred. No. 8.84e-17;
  Matches     5;  Conservative  129;  Mismatches  88;  Indels   0;  Gaps   0;

Db   1219 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1278
           ::::  : :::::::: ::::     : :  :  :: :: :    :::   :: ::::::
Qy     12 CTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTT 71

Db   1279 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1338
            :::::::::  ::::::::  :  ::: : :  . :    ::   : :  :::: :
Qy     72 GTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAA 131

Db   1339 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1398
           ::     ::       ::::::::  :   :::::::: :   :::  ::::::    ::
Qy    132 CTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTG 191

Db   1399 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440
           :::      ::: :::::::::::: :::::::::  | ||||
Qy    192 CTTGAAAANTCTGCCTCCTCCTCCATCTCCCTTCAGGGACCA 233


RESULT    2
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC    INFORMATION FOR SEQ ID NO:  5:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  215 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  protein
CC      FEATURE:
CC        NAME/KEY:  misc_feature
CC        LOCATION:  1..215
CC        OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC        OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           15.1%;  Score 36;  DB 1;  Length 215;
  Best Local Similarity  14.7%;  Pred. No. 1.01e-09;
  Matches    28;  Conservative   77;  Mismatches  83;  Indels   2;  Gaps   2;

Db       1 MTNVTMSSSSVVSRTASCNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNK 60
             :| : :::  :: : |:  ::|:::|     : |    : :|:|:  ||:::  :: : :
Cp     189 ATTCCAGGAAGCTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTC-AGT 131

Db      61 YSSANYNYGGNNVGAAKTHYYTHTNVSGADSKTVTDSYNASGTSSSNGGTD-GNRSGADS 119
             :::|   :    :   : ::: :| :   ::  :|::: |:| :::    : | ::||
Cp     130 TGCAAAGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACA 71

Db     120 YGSSKTAMTSRNRTGKTANNAVDSRNMGDASVGSDKNTKKHAKNSADGKVGSKNNGDRNN 179
             |::: |  :  :       |:::: :   ::|::  |:: |: :|:| : :   |::
Cp      70 AGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATGGAGGAGACACTGAAGT 11

Db     180 RYGTGTKSNV 189
             :  ||:: :
Cp      10 CCCGGTGGCC 1


RESULT    3
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
CC        TELEFAX:  212-838-3884
CC    INFORMATION FOR SEQ ID NO:  22:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  965 base pairs
CC        TYPE:  nucleic acid
```

```
CC        STRANDEDNESS:  unknown
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

  Query Match           13.0%;  Score 31;  DB 3;  Length 965;
  Best Local Similarity  19.3%;  Pred. No. 1.18e-06;
  Matches    23;  Conservative   53;  Mismatches  42;  Indels   1;  Gaps   1;

Db    845 SVTAADTAVYYCVRGRSYDSDGGDYWGGTTVTVSSHUVKDMTSSSSASVGDRVTTCRSST 904
          ::|  :| ::  ::| ::::::|| ::   |: :: ::::::|::::   |: :| | : |
Qy     23 CCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTT 82

Db    905 THGNGNTYYWYKGKAKYRVSNRSGVSRSGSGSGTDYTTSSDATYYCGTHARTGGTKVKG 963
          :    |:::::| |:: :  : |: :: :|:  :  |  :| :: | :| ||  :::|
Qy     83 GACC-CTCCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG 140


RESULT    4
ID   US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.
AC   xxxxxx
DT
DE   Sequence 5, Application US/08238163
CC   Sequence 5, Application US/08238163
CC   Patent No. 5569830
CC    GENERAL INFORMATION:
CC      APPLICANT:  BENNETT, Alan
CC      APPLICANT:  LABAVITCH, John M.
CC      APPLICANT:  POWELL, Ann
CC      APPLICANT:  STOTZ, Henrik
CC      TITLE OF INVENTION:  PLANT INHIBITORS OF FUNGAL
CC      TITLE OF INVENTION:  POLYGALACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE
CC      NUMBER OF SEQUENCES:  24
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Townsend and Townsend Khourie and Crew
CC        STREET:  Steuart Street Tower, One Market Plaza
CC        CITY:  San Francisco
CC        STATE:  California
CC        COUNTRY:  US
CC        ZIP:  94105-1493
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/238,163
CC        FILING DATE:  03-MAY-1994
CC        CLASSIFICATION:  800
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Bastian, Kevin L.
CC        REGISTRATION NUMBER:  34,774
CC        REFERENCE/DOCKET NUMBER:  2307E-540
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (415) 543-9600
CC        TELEFAX:  (415) 543-5043
CC    INFORMATION FOR SEQ ID NO:  5:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  215 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  protein
CC      FEATURE:
CC        NAME/KEY:  misc_feature
CC        LOCATION:  1..215
CC        OTHER INFORMATION:  /standard_name= "Deduced amino acid
CC        OTHER INFORMATION:  sequence of PGIP from bean."
SQ   SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

  Query Match           11.3%;  Score 27;  DB 1;  Length 215;
  Best Local Similarity  20.0%;  Pred. No. 2.69e-04;
  Matches    22;  Conservative   42;  Mismatches  45;  Indels   1;  Gaps   1;

Db     18 CNDKAKKDGNTTSSWTTDCCNRTWGVCDTDTTYRVNNDSGHNKYSSANYNYGGNNVGAAK 77
          | :: : :| |  ::| :   :  |:|: :| ::: : |:  : :|   :|   :|| :
Qy     89 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTG-AAGGAGT 147

Db     78 THYYTHTNVSGADSKTVTDSYNASGTSSSNGGTDGNRSGADSYGSSKTAM 127
             :::|:| :: |: : :| ::  :  :: | |:   :||: :|: : |:
Qy    148 CTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAA 197


RESULT    5
ID   US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.
AC   xxxxxx
DT
DE   Sequence 22, Application US/08388672A
CC   Sequence 22, Application US/08388672A
CC   Patent No. 5795961
CC    GENERAL INFORMATION:
CC      APPLICANT:  Wallace, T. Paul
CC      APPLICANT:  Harris, William J.
CC      APPLICANT:  Carr, Frank J.
CC      APPLICANT:  Old, Lloyd J.
CC      APPLICANT:  Welt, Sydney
CC      APPLICANT:  Kitamura, Kunio
CC      TITLE OF INVENTION:  Recombinant Human Anti-Lewis B
CC      TITLE OF INVENTION:  Antibodies
CC      NUMBER OF SEQUENCES:  25
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Felfe and Lynch
CC        STREET:  805 Third Avenue
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10022
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/388,672A
CC        FILING DATE:  14-FEB-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Hanson, No. 5795961man D.
CC        REGISTRATION NUMBER:  30,946
CC        REFERENCE/DOCKET NUMBER:  LUD 5409
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212-688-9200
CC        TELEFAX:  212-838-3884
CC    INFORMATION FOR SEQ ID NO:  22:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  965 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  unknown
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

  Query Match           10.9%;  Score 26;  DB 3;  Length 965;
  Best Local Similarity  17.8%;  Pred. No. 1.01e-03;
  Matches    16;  Conservative   43;  Mismatches  30;  Indels   1;  Gaps   1;

Db    781 GURHUVHVSGGVRSTSTCTASDYTTSYWGWVRGRGWGDYGGGYTNYNGKRGRVTMADTSS 840
          |: ::::::||: :|:| | ::: |:   ::  :  |: || :    |::|:: :|: ::
Cp    124 GTCATACCCGGCCCTGT-TCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 66

Db    841 NSRSSVTAADTAVYYCVRGRSYDSDGGDYW 870
            ::::  |   : : : |:::::: :::
Cp     65 CAGCCCAGACCCCATAGTGGCCACTGCGCT 36


RESULT    6
```

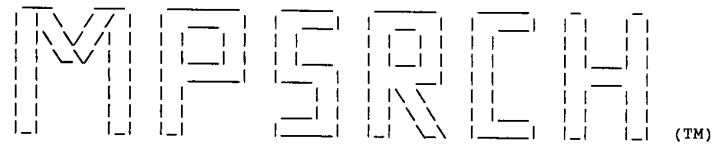
```
ID   PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.
AC   xxxxxx
DT
DE   Sequence 99, Application PC/TUS9511934
CC   Sequence 99, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  99:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  75 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 75 BP; 1 A; 1 C; 7 G; 5 T; 61 OTHER.

  Query Match            9.2%;  Score 22;  DB 4;  Length 75;
  Best Local Similarity   8.7%;  Pred. No. 1.67e-01;
  Matches     6;  Conservative   19;  Mismatches  44;  Indels   0;  Gaps   0;

Db       5 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 64
           |   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :     :  :  :
Qy      54 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 113

Db      65 NBGGTTGTG 73
             :|| | ||
Qy     114 CCGGGTATG 122


RESULT    7
ID   PCT-US95-11934-98 STANDARD; DNA; UNC; 81 BP.
AC   xxxxxx
DT
DE   Sequence 98, Application PC/TUS9511934
CC   Sequence 98, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC    INFORMATION FOR SEQ ID NO:  98:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  81 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 81 BP; 6 A; 6 C; 4 G; 5 T; 60 OTHER.

  Query Match            9.2%;  Score 22;  DB 4;  Length 81;
  Best Local Similarity   7.6%;  Pred. No. 1.67e-01;
  Matches     5;  Conservative   20;  Mismatches  41;  Indels   0;  Gaps   0;

Db       5 TGTCTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNV 64
           || | ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Cp      96 TGCCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGC 37

Db      65 NNVNNV 70
             :  :
Cp      36 TCCTGG 31


RESULT    8
ID   PCT-US95-11934-97 STANDARD; DNA; UNC; 82 BP.
AC   xxxxxx
DT
DE   Sequence 97, Application PC/TUS9511934
CC   Sequence 97, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
```

```
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  97:
CC       SEQUENCE CHARACTERISTICS:
CC        LENGTH:  82 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC       MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 82 BP; 1 A; 2 C; 10 G; 8 T; 61 OTHER.

  Query Match             9.2%;  Score 22;  DB 4;  Length 82;
  Best Local Similarity   8.7%;  Pred. No. 1.67e-01;
  Matches     6;  Conservative   19;  Mismatches  44;  Indels   0;  Gaps   0;

Db      12 GNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBN 71
           |   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :     :  :  :
Qy      54 GGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGG 113

Db      72 NBGGTTGTG 80
            :|| | ||
Qy     114 CCGGGTATG 122


RESULT    9
ID   US-08-418-444A-1 STANDARD; DNA; UNC; 3088 BP.
AC   xxxxxx
DT
DE   Sequence 1, Application US/08418444A
CC   Sequence 1, Application US/08418444A
CC   Patent No. 5773688
CC    GENERAL INFORMATION:
CC      APPLICANT:  KURODA, HISAO
CC      APPLICANT:  HIROTA, NAOHIKO
CC      APPLICANT:  ITO, KAZUTOSHI
CC      TITLE OF INVENTION:  GENE EXPRESSION VECTOR USING THE GENE
CC      TITLE OF INVENTION:  EXPRESSION REGULATING REGION OF THE ADP RIBOSYLATION
CC      TITLE OF INVENTION:  FACTOR
CC      NUMBER OF SEQUENCES:  9
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
CC        STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC        CITY:  ARLINGTON
CC        STATE:  VIRGINIA
CC        COUNTRY:  USA
CC        ZIP:  22202
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/418,444A
CC        FILING DATE:  07-APR-1995
CC        CLASSIFICATION:  800
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  JP HEI 6-71048
CC        FILING DATE:  08-APR-1994
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Oblon, No. 5773688man F.
CC        REGISTRATION NUMBER:  24,618
CC        REFERENCE/DOCKET NUMBER:  2589-024-0
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703) 413-3000
CC        TELEFAX:  (703) 413-2220
CC        TELEX:  248855 OPAT UR
CC     INFORMATION FOR SEQ ID NO:  1:
CC       SEQUENCE CHARACTERISTICS:
CC        LENGTH:  3088 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  double
CC        TOPOLOGY:  linear
CC       MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 3088 BP; 716 A; 761 C; 672 G; 939 T; 0 OTHER.

  Query Match             9.2%;  Score 22;  DB 3;  Length 3088;
  Best Local Similarity  72.0%;  Pred. No. 1.67e-01;
  Matches    36;  Conservative    0;  Mismatches  14;  Indels   0;  Gaps   0;

Db      15 CCGAAGATGGGCCAGAGGGGGGCCAGGGGCTGCCCAGGCGCCCTGCTGGC 64
           || | || || | ||||| || | |||||| |||||| | || |  ||||
Cp      94 CCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGC 45


RESULT   10
ID   US-08-471-052A-145 STANDARD; DNA; UNC; 65 BP.
AC   xxxxxx
DT
DE   Sequence 145, Application US/08471052A
CC   Sequence 145, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  145:
CC       SEQUENCE CHARACTERISTICS:
CC        LENGTH:  65 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC       MOLECULE TYPE:  DNA
SQ   SEQUENCE 65 BP; 3 A; 3 C; 3 G; 2 T; 54 OTHER.

  Query Match             8.8%;  Score 21;  DB 1;  Length 65;
  Best Local Similarity  13.8%;  Pred. No. 5.69e-01;
  Matches     9;  Conservative   16;  Mismatches  40;  Indels   0;  Gaps   0;

Db       1 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNA 60
           || ||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  |
Cp     178 CTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATA 119

Db      61 CCTGG 65
           || ||
Cp     118 CCCGG 114


RESULT   11
ID   US-08-471-052A-144 STANDARD; DNA; UNC; 66 BP.
AC   xxxxxx
DT
```

```
DE   Sequence 144, Application US/08471052A
CC   Sequence 144, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  144:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  66 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 66 BP; 2 A; 3 C; 4 G; 2 T; 55 OTHER.

  Query Match             8.8%;  Score 21;  DB 1;  Length 66;
  Best Local Similarity   5.2%;  Pred. No. 5.69e-01;
  Matches     3;  Conservative   18;  Mismatches  37;  Indels   0;  Gaps   0;

Db      4 AGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBC 61
          ||   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy     41 AGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGC 98


RESULT   12
ID   US-08-471-052A-143 STANDARD; DNA; UNC; 68 BP.
AC   xxxxxx
DT
DE   Sequence 143, Application US/08471052A
CC   Sequence 143, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  143:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  68 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 68 BP; 3 A; 3 C; 5 G; 3 T; 54 OTHER.

  Query Match             8.8%;  Score 21;  DB 1;  Length 68;
  Best Local Similarity  13.8%;  Pred. No. 5.69e-01;
  Matches     9;  Conservative   16;  Mismatches  40;  Indels   0;  Gaps   0;

Db      4 CTAGAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNA 63
          || ||:  :  :  :  :     :  :  :  :  :  :  :  :     :  :  :  |
Cp    178 CTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATA 119

Db     64 CCTGG 68
          || ||
Cp    118 CCCGG 114


RESULT   13
ID   US-08-471-052A-142 STANDARD; DNA; UNC; 69 BP.
AC   xxxxxx
DT
DE   Sequence 142, Application US/08471052A
CC   Sequence 142, Application US/08471052A
CC   Patent No. 5625033
CC    GENERAL INFORMATION:
CC      APPLICANT:  Kay, B. K.
CC      APPLICANT:  Fowlkes, D. M.
CC      TITLE OF INVENTION:  Totally Synthetic Affinity Reagents
CC      NUMBER OF SEQUENCES:  166
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  U.S.A.
CC        ZIP:  10036-2711
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/471,052A
CC        FILING DATE:  06-JUNE-1995
CC        CLASSIFICATION:  530
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-179
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  212 790-9090
CC        TELEFAX:  212 869-8864/9741
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  142:
CC      SEQUENCE CHARACTERISTICS:
```

```
CC        LENGTH:  69 bases
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  unknown
CC      MOLECULE TYPE:  DNA
SQ   SEQUENCE 69 BP; 2 A; 4 C; 6 G; 2 T; 55 OTHER.

  Query Match             8.8%;  Score 21;  DB 1;  Length 69;
  Best Local Similarity   5.2%;  Pred. No. 5.69e-01;
  Matches     3;  Conservative   18;  Mismatches  37;  Indels   0;  Gaps   0;

Db       7 AGNNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBC 64
           ||   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy      41 AGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGC 98


RESULT   14
ID   PCT-US95-11934-94 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 94, Application PC/TUS9511934
CC   Sequence 94, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  94:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 3 A; 4 C; 3 G; 1 T; 63 OTHER.

  Query Match             8.8%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity   1.7%;  Pred. No. 5.69e-01;
  Matches     1;  Conservative   20;  Mismatches  39;  Indels   0;  Gaps   0;

Db       9 BNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNNBNA 68
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :|
Qy      37 GCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCA 96


RESULT   15
ID   PCT-US95-11934-100 STANDARD; DNA; UNC; 74 BP.
AC   xxxxxx
DT
DE   Sequence 100, Application PC/TUS9511934
CC   Sequence 100, Application PC/TUS9511934
CC    GENERAL INFORMATION:
CC      APPLICANT:  Cytogen Corporation
CC      TITLE OF INVENTION:  Antigen Binding Peptides (Abtides) From
CC      TITLE OF INVENTION:  Peptide Libraries
CC      NUMBER OF SEQUENCES:  103
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  Pennie & Edmonds
CC        STREET:  1155 Avenue of the Americas
CC        CITY:  New York
CC        STATE:  New York
CC        COUNTRY:  USA
CC        ZIP:  10036
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  PCT/US95/11934
CC        FILING DATE:  20-SEP-1995
CC        CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Misrock, S. Leslie
CC        REGISTRATION NUMBER:  18,872
CC        REFERENCE/DOCKET NUMBER:  1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (212) 790-9090
CC        TELEFAX:  (212) 869-9741/8864
CC        TELEX:  66141 PENNIE
CC     INFORMATION FOR SEQ ID NO:  100:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  74 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  single
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 74 BP; 6 A; 6 C; 1 G; 1 T; 60 OTHER.

  Query Match             8.8%;  Score 21;  DB 4;  Length 74;
  Best Local Similarity   7.5%;  Pred. No. 5.69e-01;
  Matches     5;  Conservative   19;  Mismatches  43;  Indels   0;  Gaps   0;

Db       5 AVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNN 64
           |:  :  :  :  :  :  :     :  :  :  :  :  :  :  :  :  :  :
Cp     125 AGTCATACCCGGCCCTGTTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGG 66

Db      65 NACCACA 71
             | | ||
Cp      65 CAGCCCA 59


Search completed: Sun Oct 24 16:40:52 1999
Job time : 23 secs.
```

****



****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Oct 24 16:30:43 1999; MasPar time 462.56 Seconds
1210.689 Million cell updates/sec

Tabular output not generated.

Title: >US-09-092-296-1
Description: (1-239) from US09092296.seq
Perfect Score: 238
N.A. Sequence: 1 GGCCACCGGGACTTCAGTGT.....CCCTTCAGGGACCAGCGTCA 239
Comp: CCGGTGGCCCTGAAGTCACA.....GGGAAGTCCCTGGTCGCAGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est2 20:gb_est20
21:gb_est21 22:gb_est22 23:gb_est23 24:gb_est24
25:gb_est25 26:gb_est26 27:gb_est27 28:gb_est28
29:gb_est29 30:gb_est3 31:gb_est4 32:gb_est5 33:gb_est6
34:gb_est7 35:gb_est8 36:gb_est9 37:gb_gss1 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 9.792; Variance 2.100; scale 4.663

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	57	23.9	252	17	AA754459	97SN1787 Rice Immature	1.50e-56
c	2	54	22.7	328	23	AI136523	UI-R-C2p-nq-e-02-0-UI.	1.62e-51
c	3	51	21.4	252	17	AA754459	97SN1787 Rice Immature	1.52e-46
c	4	39	16.4	247	17	AA754458	97SN1784 Rice Immature	2.04e-27
	5	33	13.9	247	17	AA754458	97SN1784 Rice Immature	1.76e-18
c	6	31	13.0	2275	20	AF034173	AF034173 Human mRNA (T	1.24e-15
	7	27	11.3	2275	20	AF034173	AF034173 Human mRNA (T	3.49e-10
	8	24	10.1	348	19	F06958	HSC1QC101 normalized i	2.31e-06
	9	23	9.7	238	12	AA376266	EST88915 HSC172 cells	3.76e-05
	10	23	9.7	311	11	AA323964	EST26816 Cerebellum II	3.76e-05
	11	23	9.7	364	34	W33870	mc56b03.r1 Soares mous	3.76e-05
c	12	23	9.7	412	18	AA769782	ah71b05.s1 Soares_test	3.76e-05
	13	23	9.7	426	8	T15255	crs855 lambdaZAPST Ric	3.76e-05
	14	23	9.7	481	14	AA498488	vh44b06.r1 Barstead mo	3.76e-05
	15	23	9.7	1287	20	AF038250	AF038250 Human mRNA (T	3.76e-05
	16	22	9.2	308	8	D40392	RICS2342A Rice shoot O	5.65e-04
	17	22	9.2	339	19	F08745	HSC1DB011 normalized i	5.65e-04
c	18	22	9.2	404	17	AA730403	nw42c07.s1 NCI_CGAP_Ew	5.65e-04
	19	22	9.2	433	34	W79098	zd75h10.r1 Soares_feta	5.65e-04
	20	22	9.2	441	14	C28493	C28493 Rice callus cDN	5.65e-04
c	21	22	9.2	634	22	AI063013	GH02423.5prime GH Dros	5.65e-04
	22	22	9.2	1025	37	B12587	F22O11-Sp6.1 IGF Arabi	5.65e-04
	23	21	8.8	288	11	AA335414	EST39832 Epididymus Ho	7.76e-03
c	24	21	8.8	299	8	M79528	wEST00065 Mixed stage,	7.76e-03
	25	21	8.8	317	19	F10052	HSC39H122 normalized i	7.76e-03
	26	21	8.8	343	8	T47417	yb13f12.r1 Stratagene	7.76e-03
	27	21	8.8	384	23	AI160316	qb66d04.x1 Soares_feta	7.76e-03
c	28	21	8.8	395	41	AQ375965	RPCI11-161H11.TV RPCI1	7.76e-03
c	29	21	8.8	401	42	AQ444216	GSSTc0313 Trypanosoma	7.76e-03
	30	21	8.8	424	33	W13114	ma86e10.r1 Soares mous	7.76e-03
c	31	21	8.8	429	28	AI509530	vx09h04.y1 Soares 2NbM	7.76e-03
c	32	21	8.8	440	18	AA822673	vw14f01.r1 Soares 2NbM	7.76e-03
	33	21	8.8	468	35	AA046996	zf50a10.r1 Soares reti	7.76e-03
	34	21	8.8	479	26	AI398157	NCSP4B8T7 Subtracted P	7.76e-03
	35	21	8.8	499	37	B45020	HS-1060-B1-B01-MF.abi	7.76e-03
c	36	21	8.8	516	31	H41536	yp71c01.s1 Soares adul	7.76e-03
c	37	21	8.8	524	10	AA239967	mw24g07.r1 Soares mous	7.76e-03
	38	21	8.8	550	20	AA898701	NCM6G7T7 Mycelial Neur	7.76e-03
c	39	21	8.8	567	30	R61539	yh16f01.s1 Soares infa	7.76e-03
c	40	21	8.8	601	32	N37023	yy40d03.s1 Soares mela	7.76e-03
c	41	21	8.8	688	37	AG015068	Homo sapiens genomic D	7.76e-03
c	42	21	8.8	694	36	AA140679	CK00462.5prime CK Dros	7.76e-03
	43	21	8.8	749	37	B01542	cSRL-133c3-u cSRL flow	7.76e-03
	44	21	8.8	762	37	B19344	T27I1-Sp6.1 TAMU Arabi	7.76e-03
c	45	21	8.8	796	37	AF035986	Salmonella typhimurium	7.76e-03

ALIGNMENTS

RESULT 1
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID g2801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
FEATURES Location/Qualifiers
source 1. .252
/organism="Oryza sativa"

/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5'end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 23.9%; Score 57; DB 17; Length 252;
Best Local Similarity 11.4%; Pred. No. 1.50e-56;
Matches 24; Conservative 109; Mismatches 75; Indels 3; Gaps 3;

```
Db     23 YBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWBV 82
          ::|:  :: : :|::| :: ::   ::|||: : :    : :|:|||: ::  : : :: :
Qy     29 TCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCT 88

Db     83 BNTKVDVGNHTRCSRWRBVTRMAH-YHDYTNCBBYNNNDYHMWHBBMYBBTGCMTCTMWC 141
          : |:::: :|  :::::: :  :::   ::    : ::::::: ::|| :    :|
Qy     89 CCTTGGCAGCTCACATGGAACAGGGCCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTC 148

Db    142 WBHYNTKCTASGWHTSTNYDVKS-STNTWGVTBSYDKSMHGYWCSBBVKYHTKV-STTRA 199
          :::: |:  |  :: :| ::::: :|  :|:|:::::::: : | :::  : :: :
Qy    149 TTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAATTGCTTGAAAANTCTGCCTC 208

Db    200 TRSYTCVRKYCVMWMTKKVVKKYHVVBBGCH 230
            ::||:::: ::  |  ::: :::::: |:
Qy    209 CTCCTCCATCTCCCTTCAGGGACCAGCGTCA 239
```

RESULT 2
LOCUS AI136523 328 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C2p-nq-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone UI-R-C2p-nq-e-02-0-UI 3', mRNA sequence.
ACCESSION AI136523
NID g3637300
VERSION AI136523.1 GI:3637300
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 328)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 14, 1998 this sequence version replaced gi:1877567.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.

FEATURES Location/Qualifiers
source 1. .328
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
/db_xref="taxon:10116"
/clone="UI-R-C2p-nq-e-02-0-UI"
/clone_lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

BASE COUNT 62 a 77 c 98 g 91 t
ORIGIN

Query Match 22.7%; Score 54; DB 23; Length 328;
Best Local Similarity 74.2%; Pred. No. 1.62e-51;
Matches 89; Conservative 0; Mismatches 30; Indels 1; Gaps 1;

```
Db    181 TGACATTGGTCCCCGGTGAGAGGTGGAGGAGGAGGCAGATCTTTTGGAGCATGTCCAGAA 240
          ||||  ||||||| |  | ||| ||||||||||||||||  ||||  ||||  ||||| |
Cp    239 TGACGCTGGTCCCTGAAGGGAGATGGAGGAGGAGGCAGAN-TTTTCAAGCAATTCCAGGA 181

Db    241 AGCCGGAGTCTTGGGAGGTCTTGGCCTGAGATGCCTCGGTCAGTTTTACTTGCAAAGTCA 300
          ||| ||| || | |||||  || | | || | | |||  |||| || | ||||||||||| .
Cp    180 AGCTGGACTCATAGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCA 121
```

RESULT 3
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID g2801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr

```
          Seq primer: M13 Reverse Primer.
FEATURES             Location/Qualifiers
     source          1. .252
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
                     /db_xref="taxon:4530"
                     /map="6"
                     /clone="97SN1787"
                     /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
BASE COUNT        5 a     21 c     12 g     35 t    179 others
ORIGIN

  Query Match           21.4%;  Score 51;  DB 17;  Length 252;
  Best Local Similarity   9.8%;  Pred. No. 1.52e-46;
  Matches    21;  Conservative  108;  Mismatches  83;  Indels   2;  Gaps   2;

Db      22 SYBCHGNBVWVCVASHGNYMSVHNCTBRGTHCDCKNVNWSTMTWGTVNWBNVSGDWHYWB 81
           ::: :  ::::   ::|   ::  | :: |: :|  : : |      : ::  :|: :::
Cp     228 CCTGAAGGGAGATGGAGGAGGAGGCAGANTTTCAAGCAATTCCAGGAAGCTGGACTCAT 169

Db      82 VBNTKVDVGNHTRCSRWRBVTRMAHYHDYTNCBBYNNNDYHMWHBBMYBBTGCMTCTMWC 141
           ::  ::::  :  | ::: : :: :::::  |:::   : :::::::: :   |  | ::
Cp     168 AGGAGGAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTG 109

Db     142 WBHYNTKCTASGWHTSTNYDVKSSTNTWGVTBSYDKSMHGYW-C-SBBVKYHTKVSTTRA 199
           :::: |:  |: :  :   :::::|  :|: :  : :::|   | ::::  :  :: |:
Cp     108 TTCCATGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAG 49

Db     200 TRSYTCVRKYCVMWMTKKVVKKYHVVBBGCHBTD 233
           |:::  : :: :  :|:::::: ::: : |::|:
Cp      48 TGGCCACTGCGCTCCTGGGATGGAGGAGACACTG 15


RESULT    4
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
VERSION     AA754458.1  GI:2801164
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797455.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
                     /db_xref="taxon:4530"
                     /map="6"
                     /clone="97SN1784"
                     /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match           16.4%;  Score 39;  DB 17;  Length 247;
  Best Local Similarity  10.2%;  Pred. No. 2.04e-27;
  Matches    21;  Conservative  100;  Mismatches  82;  Indels   3;  Gaps   3;

Db      34 VRRVGTTTNNGKHNGRTTTWNDCSDNAHCRYTVBWYYARSKYGYGTBYYSWNVDTNTGGT 93
           ::::| |   |:: |:   : :  :  : :  : ::: :::    |:  :: ::|  |
Cp     222 GGGAGATGGAGGAGGAGGCAGANTTTTCAAGCAATTCCAGGAAGCTGGACT-CATAGGAG 164

Db      94 GVGKTTVNVHSGWNNRCSNSVVYVWBTAYCDYBHYBDRANHVDDTRCTNDRGYCNYTASD 153
           |: :||: ::: :  : : :: :::: :|: :: :::| ::::  |  :  :  :| ::
Cp     163 GAATTTGTCAGAAAAGACTCCTTCAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCA 104

Db     154 NGTSAT-KRVTGYDKTDSDCGGGCWRKVTYGSS-BYBRCGVNVMVRTTSMWTDKSTKMBS 211
            ||:|  :::   :: :::     ::::   :  : ::| : ::::  ::  : :|: ::
Cp     103 TGTGAGCTGCCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCC 44

Db     212 MDMSRRSRVHYGRWMBNKKRGMSRNW 237
           :  : :: :::|: :: : :|::: :
Cp      43 ACTGCGCTCCTGGGATGGAGGAGACA 18


RESULT    5
LOCUS       AA754458      247 bp    mRNA            EST       20-JAN-1998
DEFINITION  97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 97SN1784, mRNA sequence.
ACCESSION   AA754458
NID         g2801164
VERSION     AA754458.1  GI:2801164
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797455.

            Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
```

```
                    /db_xref="taxon:4530"
                    /map="6"
                    /clone="97SN1784"
                    /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                    /tissue_type="Immature Seed"
                    /dev_stage="5 days after pollination"
                    /lab_host="E. coli SOLR"
BASE COUNT        7 a     16 c     21 g     34 t    169 others
ORIGIN

  Query Match           13.9%;  Score 33;  DB 17;  Length 247;
  Best Local Similarity  13.3%;  Pred. No. 1.76e-18;
  Matches    26;  Conservative   88;  Mismatches  79;  Indels   3;  Gaps   3;

Db     48 GRTTTWNDCSDNAHCRYTVBWYYARS-KYGYGTBYYSWNVDTNTGGTGVGKTTVNVHSGW 106
          |: ||  : ::  :| : :: ::|:: ::|:  :  :  ::   ||  : :  : :::
Qy     10 GACTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCC 69

Db    107 NNRCSNSVVYVWBTAYCDYBHYBDRANHVDDTRCTNDRGYCNYTASDNGTSATKRVTGYD 166
          : :  :::::: |:| ::::::   ::    | :: :     :: || ||:::| ::
Qy     70 TTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGGGCCGGGT-ATGACTTTG 128

Db    167 KTDSDCGGGCWRKVTYGS-SBYBRCGVNVMVRTTSMWTDKSTKMBSMDMSRRSRVHYGRW 225
          :::   ||:: :   : :::: | : ::::||:::  : |::: : ::    :: |::
Qy    129 CAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTATGAGTCCAGCTTCCTGGAA 188

Db    226 MBNKKRGMSRNWTDTK 241
          :  : |: :  | |:
Qy    189 TTGCTTGAAAANTCTG 204


RESULT    6
LOCUS       AF034173     2275 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
VERSION     AF034173.1  GI:2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045115.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match           13.0%;  Score 31;  DB 20;  Length 2275;
  Best Local Similarity   9.6%;  Pred. No. 1.24e-15;
  Matches    11;  Conservative   64;  Mismatches  39;  Indels   1;  Gaps   1;

Db   1480 RYKRWKRRKGRRKRMTGMYKRMYRAMMMAMCAMMACWWYYWKMRGMKKCWKYRKYKKYTS 1539
          ::::::::::|::: :   ::  ::| ::|   ::  :: ::: :     :: ::   :|
Cp    217 ATGGAGGAGGAGGCAGANTTTTCAAGCAATTCCAGGAAGCTGGACTCATAGGAGGAATTT 158

Db   1540 TYYKSWSRWYWYTTYTYWYCWCCTSMKSASCAMMRWMGYMGSRSSRSYWGYWGSM 1594
          :: :: :: ::| :|: :  | |:::  :||::::: :: : :: :::|:: ::
Cp    157 GTCAGAAAAGACTCCTT-CAGCTTCAGTTGCAAAGTCATACCCGGCCCTGTTCCA 104


RESULT    7
LOCUS       AF034173     2275 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION   AF034173
NID         g2707735
VERSION     AF034173.1  GI:2707735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2275)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045115.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .2275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon2 contig"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      438 a    619 c    470 g    599 t    149 others
ORIGIN

  Query Match           11.3%;  Score 27;  DB 20;  Length 2275;
  Best Local Similarity  16.2%;  Pred. No. 3.49e-10;
  Matches    11;  Conservative   39;  Mismatches  17;  Indels   1;  Gaps   1;

Db   1556 YWYCWCCTSMKSASCAMMRWMGYMGS-RSSRSYWGYWGSMSGCYGMTKRYYRYSWTGWTK 1614
          :::|: | ::::|:| :::: |:: : : ::: : : |:: ||: :|:::: :: | :|:
Qy     24 CTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTCCTCTTG 83

Db   1615 WTTWWYMW 1622
          :  ::::
Qy     84 ACCCTCCT 91


RESULT    8
LOCUS       F06958       348 bp    mRNA            EST       20-FEB-1995
DEFINITION  HSC1QC101 normalized infant brain cDNA Homo sapiens cDNA clone
            c-1qc10, mRNA sequence.
ACCESSION   F06958
NID         g672595
VERSION     F06958.1  GI:672595
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 348)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
```

```
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     On Sep 21, 1992 this sequence version replaced gi:279286.

            Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-1qc10
            Insert Length: 639   Std Error: 0.00
            Seq primer: (-21)M13_universal
            High quality sequence stop: 150.
FEATURES             Location/Qualifiers
     source          1. .348
                     /organism="Homo sapiens"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total
                     brain;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
                     /db_xref="taxon:9606"
                     /clone="c-1qc10"
                     /clone_lib="normalized infant brain cDNA"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
BASE COUNT       84 a     91 c     90 g     81 t      2 others
ORIGIN

  Query Match           10.1%;  Score 24;  DB 19;  Length 348;
  Best Local Similarity  67.2%;  Pred. No. 2.31e-06;
  Matches    45;  Conservative    0;  Mismatches  22;  Indels   0;  Gaps   0;

Db      5 TGTTGCCTCCCTCCCAGGATCCCTTTGGTGAGTATGGTGTTCAGGNTGCACCACCACCAC 64
          |||  ||||| |||||||| | |  |||  | ||||| ||   || ||| ||    || |
Qy     18 TGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTC 77

Db     65 CTCTAGA 71
          |||| ||
Qy     78 CTCTTGA 84


RESULT    9
LOCUS       AA376266      238 bp    mRNA            EST       21-APR-1997
DEFINITION  EST88915 HSC172 cells II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA376266
NID         g2028809
VERSION     AA376266.1  GI:2028809
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC191210
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .238
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
                     /db_xref="ATCC (inhost):180944"
                     /db_xref="taxon:9606"
                     /clone_lib="HSC172 cells II"
                     /cell_type="fibroblast"
                     /cell_line="HSC172 (60PDL)"
                     /dev_stage="fetal"
BASE COUNT       48 a     70 c     50 g     66 t      4 others
ORIGIN

  Query Match           9.7%;  Score 23;  DB 12;  Length 238;
  Best Local Similarity  67.2%;  Pred. No. 3.76e-05;
  Matches    45;  Conservative    0;  Mismatches  22;  Indels   0;  Gaps   0;

Db    141 TGTTGCCTCCCTCCCAGGATCCCTTTGGTGAGTATGGTGTTCAGGATGCACCACCACCAC 200
          |||  ||||| |||||||| | |  |||  | ||||| ||   || ||| ||    || |
Qy     18 TGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCCTTGTCCTC 77

Db    201 CTCTAGA 207
          |||| ||
Qy     78 CTCTTGA 84


RESULT   10
LOCUS       AA323964      311 bp    mRNA            EST       20-APR-1997
DEFINITION  EST26816 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA323964
NID         g1976290
VERSION     AA323964.1  GI:1976290
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
```

Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT On Apr 14, 1993 this sequence version replaced gi:693635.

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
source 1. .311
/organism="Homo sapiens"
/note="Organ: brain; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
/db_xref="ATCC (inhost):124516"
/db_xref="taxon:9606"
/clone_lib="Cerebellum II"
/tissue_type="cerebellum"
/dev_stage="adult"

BASE COUNT 59 a 110 c 74 g 64 t 4 others
ORIGIN

Query Match 9.7%; Score 23; DB 11; Length 311;
Best Local Similarity 89.7%; Pred. No. 3.76e-05;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 8 TGGCCACTGGGGGGCCTGGGCTGCCCCTT 36
|||||||| |||| ||||||||||||||
Qy 43 TGGCCACTATGGGGTCTGGGCTGCCCCTT 71

RESULT 11
LOCUS W33870 364 bp mRNA EST 13-MAY-1996
DEFINITION mc56b03.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA clone IMAGE:352493 5', mRNA sequence.
ACCESSION W33870
NID g1315794
VERSION W33870.1 GI:1315794
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 364)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 18, 1995 this sequence version replaced gi:811194.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:224293
Seq primer: ETPrimer
High quality sequence stop: 360.

FEATURES Location/Qualifiers
source 1. .364
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "
/db_xref="taxon:10090"
/clone="IMAGE:352493"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"

BASE COUNT 100 a 73 c 87 g 104 t
ORIGIN

Query Match 9.7%; Score 23; DB 34; Length 364;
Best Local Similarity 79.5%; Pred. No. 3.76e-05;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 282 AACTGAAAGGGAAGGAGGCTTTTCCCACATATTTCTCCT 320
||||||| ||||||| |||||| ||| ||| |||||
Qy 130 AACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCT 168

RESULT 12
LOCUS AA769782 412 bp mRNA EST 29-DEC-1998
DEFINITION ah71b05.s1 Soares_testis_NHT Homo sapiens cDNA clone 1321041 3', mRNA sequence.
ACCESSION AA769782
NID g2821020
VERSION AA769782.1 GI:2821020
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 412)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 9, 1998 this sequence version replaced gi:936203.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 691 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 402.

FEATURES Location/Qualifiers

```
    source          1. .412
                    /organism="Homo sapiens"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from mRNA obtained from Clontech
                    Laboratories, Inc., and primed with  a Not I - oligo(dT)
                    primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                    Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization to Cot5, and was
                    constructed  by Bento Soares and M. Fatima Bonaldo. "
                    /db_xref="taxon:9606"
                    /clone="1321041"
                    /clone_lib="Soares_testis_NHT"
                    /sex="male"
                    /lab_host="DH10B"
BASE COUNT      111 a     78 c    116 g    107 t
ORIGIN

  Query Match            9.7%;  Score 23;  DB 18;  Length 412;
  Best Local Similarity  67.2%;  Pred. No. 3.76e-05;
  Matches    45;  Conservative    0;  Mismatches  22;  Indels   0;  Gaps   0;

Db    262 TCTAGAGGTGGTGGTGGTGCATCCTGAACACCATACTCACCAAAGGGATCCTGGGAGGGA 321
          || ||||| ||    || ||| ||   || ||||| |  |||  | | |||||||| |||
Cp     84 TCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGCCACTGCGCTCCTGGGATGGA 25

Db    322 GGCAACA 328
          ||  |||
Cp     24 GGAGACA 18


RESULT   13
LOCUS       T15255        426 bp    mRNA            EST       28-JUL-1995
DEFINITION  crs855 lambdaZAPST Ricinus communis cDNA clone pcrs855, mRNA
            sequence.
ACCESSION   T15255
NID         g688907
VERSION     T15255.1  GI:688907
KEYWORDS    EST.
SOURCE      castor bean.
  ORGANISM  Ricinus communis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Euphorbiales; Euphorbiaceae; Ricinus.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   vandeLoo,F.J., Turner,S. and Somerville,C.
  TITLE     Expressed sequence tags from developing castor seeds
  JOURNAL   Plant Physiol. 108, 1141-1150 (1995)
COMMENT
            Contact: Somerville CR
            Carnegie Institution
            Carnegie Institution, 290 Panama St, Stanford, CA 94305
            Tel: 4153251521
            Email: crs@andrew.stanford.edu
            Seq primer: T3.
FEATURES              Location/Qualifiers
     source           1. .426
                      /organism="Ricinus communis"
                      /strain="Baker 296"
                      /note="Vector: lambdaZAPII; Site_1: EcoR1; Site_2: XhoI;
                      Poly(A)+ RNA was purified from developing stage III to
                      stage V (Greenwood & Bewley, Can. J. Bot. 60:1751-1760,
                      1982) endosperm plus embryo of immature castor fruits.
                      cDNA was synthesized and cloned into lambdaZAPII according
                      to the instructions of the manufacturer (Stratagene):
                      synthesis was primed from the poly(A) tail, and cloned
                      directionally into XhoI (3') and EcoRI (5') sites.  In few
                      cases, sequence data indicated that this directionality
                      was reversed.  Partial cDNA clones predominate. "
                      /db_xref="taxon:3988"
                      /clone="pcrs855"
                      /clone_lib="lambdaZAPST"
BASE COUNT      126 a     61 c     99 g    129 t     11 others
ORIGIN

  Query Match            9.7%;  Score 23;  DB 8;  Length 426;
  Best Local Similarity  71.4%;  Pred. No. 3.76e-05;
  Matches    35;  Conservative    0;  Mismatches  14;  Indels   0;  Gaps   0;

Db     77 TGACCATTCAAGTGAAGCTGAAGATGACTTGTCTGANTGATTGGNCCTA 125
          ||||  | ||| ||||||||||  | ||| |||||   |||   ||||
Qy    121 TGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCCTA 169


RESULT   14
LOCUS       AA498488      481 bp    mRNA            EST       01-JUL-1997
DEFINITION  vh44b06.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:889811 5' similar to gb:M30514 Mouse muscle nicotinic
            acetylcholine receptor gamma-subunit (MOUSE);, mRNA sequence.
ACCESSION   AA498488
NID         g2233511
VERSION     AA498488.1  GI:2233511
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1405167.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:517771
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 461.
FEATURES              Location/Qualifiers
     source           1. .481
                      /organism="Mus musculus"
                      /strain="FVB/N"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      [GTTGGATTCGGTACC], digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library constructed by Bob Barstead."
                      /db_xref="taxon:10090"
                      /clone="IMAGE:889811"
                      /clone_lib="Barstead mouse pooled organs MPLRB4"
                      /sex="mixed"
                      /tissue_type="pooled organs"
                      /dev_stage="7 day"
                      /lab_host="DH10B"
BASE COUNT      120 a    131 c    124 g    106 t
ORIGIN

  Query Match            9.7%;  Score 23;  DB 14;  Length 481;
  Best Local Similarity  76.7%;  Pred. No. 3.76e-05;
```

```
  Matches    33;  Conservative    0;  Mismatches  10;  Indels   0;  Gaps   0;

Db     277 CTTCAGTGTCAACAACCTGCTTGGAGCCCAGTGGCCAATATGG 319
           ||||||||||  |  | | |  ||||| |||||||||| |||||
Qy      12 CTTCAGTGTCTCCTCCATCCCAGGAGCGCAGTGGCCACTATGG 54


RESULT   15
LOCUS       AF038250     1287 bp    mRNA            EST       30-MAR-1998
DEFINITION  AF038250 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
            clone ntcon9, mRNA sequence.
ACCESSION   AF038250
NID         g2815880
VERSION     AF038250.1  GI:2815880
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1287)
  AUTHORS   Tripodis,N. and Ragoussis,J.
  TITLE     Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045085.

            Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES             Location/Qualifiers
     source          1. .1287
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.3"
                     /clone="ntcon9"
                     /clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      349 a    219 c    293 g    361 t     65 others
ORIGIN

  Query Match            9.7%;  Score 23;  DB 20;  Length 1287;
  Best Local Similarity  25.8%;  Pred. No. 3.76e-05;
  Matches    24;  Conservative   32;  Mismatches  37;  Indels   0;  Gaps   0;

Db     381 VVVATCNAGSCCVACVTCBCDCBTGGGDTBSHBVCMCBCBANADGATBATCGKVGGVBGA 440
           ::: ||  |:| : |:  : :|:|    |::::| |: :   :| |:|   ::  : |
Qy      53 GGGGTCTGGGCTGCCCCTTGTCCTCCTCTTGACCCTCCTTGGCAGCTCACATGGAACAGG 112

Db     441 STCMTHYNCCDCKTCGSAGVTVNHHDWSMAGGA 473
           : |   :   :|:| |:| :|: :  :::||||
Qy     113 GCCGGGTATGACTTTGCAACTGAAGCTGAAGGA 145


Search completed: Sun Oct 24 16:38:34 1999
Job time : 471 secs.
```

=> d 1-4 ibib abs

L27 ANSWER 1 OF 4 CAPLUS COPYRIGHT 1999 ACS DUPLICATE 1
ACCESSION NUMBER: 1999:518247 CAPLUS
DOCUMENT NUMBER: 131:166197
TITLE: Methods for detecting **lung diseases**
INVENTOR(S): **Cohen, Maurice**; **Friedman, Paula N.**; Gordon, Julian; Hodges, Steven C.; **Klass, Michael R.**; Kratochvil, Jon D.; Roberts-Rapp, Lisa; **Russell, John C.**; **Stroupe, Steven D.**
PATENT ASSIGNEE(S): Abbott Laboratories, USA
SOURCE: U.S., 36 pp., Cont.-in-part of U.S. Ser. No. 744,211, abandoned.
CODEN: USXXAM
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 5939265	A	19990817	US 1997-964725	19971105
PRIORITY APPLN. INFO.:			US 1996-744211	19961105

AB A set of contiguous and partially overlapping RNA sequences and polypeptides encoded thereby, designated as LU103 and transcribed from lung tissue are described. A fully sequenced clone representing the longest continuous sequence of LU103 is also disclosed. These sequences are useful in detecting, diagnosing, staging, monitoring, prognosticating, preventing or treating, or detg. the predisposition of an individual to diseases and conditions of the lung such as lung cancer.

L27 ANSWER 2 OF 4 CAPLUS COPYRIGHT 1999 ACS DUPLICATE 2
ACCESSION NUMBER: 1999:8152 CAPLUS
DOCUMENT NUMBER: 130:77055
TITLE: Protein LS170 and cDNA sequences useful for detecting **diseases** of the human **lung**
INVENTOR(S): **Billing-Medel, Patricia A.**; **Cohen, Maurice**; **Colpitts, Tracey L.**; **Friedman, Paula N.**; Gordon, Julian; Granados, Edward N.; Hodges, Steven C.;

Klass, Michael R.; Kratochvil, Jon D.; Roberts-Rapp, Lisa; **Russell, John C.**; **Stroupe, Stephen D.**
PATENT ASSIGNEE(S): Abbott Laboratories, USA
SOURCE: PCT Int. Appl., 120 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9856951	A1	19981217	WO 1998-US11601	19980611

W: CA, JP
RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE
PRIORITY APPLN. INFO.: US 1997-49183 19970611
AB A set of contiguous and partially overlapping cDNA sequences and polypeptides encoded thereby, designated as LS170 and transcribed from human lung tissue, is described. These sequences are useful for the detecting, diagnosing, staging, monitoring, prognosticating, in vivo imaging, preventing or treating, or detg. the predisposition of an individual to diseases and conditions of the lung, such as lung cancer. Also provided are antibodies which specifically bind to a LS170-encoded polypeptide or protein, and agonists or inhibitors which prevent action of tissue-specific LS170 polypeptides, which mols. are useful for the therapeutic treatment of **lung diseases**, tumors, or metastases.

L27 ANSWER 3 OF 4 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-437479 [37] WPIDS
DOC. NO. NON-CPI: N1998-340776
DOC. NO. CPI: C1998-133112
TITLE: New nucleic acid for the **lung disease** marker LU105 - polypeptides, antibodies and genes, used for diagnosis, prevention, treatment of **lung disease**, specifically cancer.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): **BILLING-MEDEL, P A**; **COHEN, M**; **COLPITTS, T L**; **FRIEDMAN, P N**; GORDON, J; GRANADOS, E N; HODGES, S C; **KLASS, M R**; KRATOCHVIL, J D; ROBERTSRAPP, L; **RUSSELL, J C**; **STROUPE, S D**
PATENT ASSIGNEE(S): (ABBO) ABBOTT LAB
COUNTRY COUNT: 19
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9833926	A1	19980806	(199837)*	EN	117

RW: AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE
W: CA JP

APPLICATION DETAILS:

Searcher : Shears 308-4994

PATENT NO	KIND	APPLICATION	DATE
WO 9833926	A1	WO 1998-US1766	19980130

PRIORITY APPLN. INFO: US 1997-791710 19970131
AN 1998-437479 [37] WPIDS
AB WO 9833926 A UPAB: 19980916
A method for detecting target LU105 nucleic acid (I) comprises treating a sample with at least one LU105-specific nucleic acid (II), or its complement. (II) is at least 50% identical with 190, 244, 225, 114, 562 or 519 bp sequences given in the specification, or their fragments and complements.

Also claimed are:

(1) (I) or its fragments able to hybridise selectively to the LU105 gene and having at least 50% identity with the 190, 244, 225, 114, 562 or 519 bp sequences given above;

(2) recombinant expression systems including (I) and control sequence;

(3) cells transformed with this expression system;

(4) LU105 polypeptides (III) at least 50% identical with the 104, 26, 19, 21, 18 or 19 amino acid (aa) sequences given in the specification or their fragments;

(5) antibodies (Ab) that bind to at least one LU105 epitope present in (III);

(6) cells transformed with the 190, 244, 225, 114, 562 or 519 bp sequences described above;

(7) LU105 specific nucleic acid (II); and

(8) genes, or their fragments, that encode a protein at least 50% identical with the 104 aa sequences as in (4).

USE - LU105 is a **lung disease** marker. Cells as in (3) are used to express recombinant (III) which are used to raise Ab. Ab are used to detect the LU105 antigen, and correspondingly this antigen is used to detect specific antibodies, in usual immunoassays. (I) and (III) are used for diagnosis, staging, monitoring, prognosis, prevention, treatment (e.g. using antisense molecules, ribozymes, Ab or other antagonists) and determination of susceptibility to, **lung disease**, specifically cancer. (III) are also used to screen for specific binding agents, potentially useful therapeutically. LU105 is a marker for **lung disease** (present at high concentration, in altered form or in an unusual body compartment).

ADVANTAGE - LU105 can be detected in blood, plasma or serum in an inexpensive, non-invasive test.

Dwg.0/6

L27 ANSWER 4 OF 4 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-286957 [25] WPIDS
DOC. NO. NON-CPI: N1998-225472
DOC. NO. CPI: C1998-088988
TITLE: Lung tissue derived polynucleotide LU103 - useful to detect, diagnose, stage, monitor, prognose, prevent, treat or determine pre-disposition to **lung disease**, e.g. **lung** cancer.

Searcher : Shears 308-4994

DERWENT CLASS: B04 D16 S03
INVENTOR(S): **COHEN, M; FRIEDMAN, P N;** GORDON, J; HODGES, S C; **KLASS, M R;** KRATOCHVILL, J D; ROBERTS-RAPP, L; **RUSSELL, J C; STROUPE, S D**
PATENT ASSIGNEE(S): (ABBO) ABBOTT LAB
COUNTRY COUNT: 19
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9820143	A1	19980514	(199825)*	EN	86

RW: AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE
W: CA JP

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9820143	A1	WO 1997-US20680	19971105

PRIORITY APPLN. INFO: US 1996-744211 19961105
AN 1998-286957 [25] WPIDS
AB WO 9820143 A UPAB: 19980715
The following are claimed: (1) a method for detecting the presence of a target LU103 polynucleotide in a test sample, comprising: (a) contacting the sample with at least 1 LU103-specific polynucleotide, and (b) detecting the target LU103 polynucleotide in the test sample, where the LU103 polynucleotide has at least 50% identity to the 269, 263, 225, 507 or 519 bp nucleic acid sequence given in the specification; (2) a method for detecting LU103 mRNA in a test sample, comprising: (a) performing reverse transcription with at least 1 primer in order to produce cDNA; (b) amplifying the cDNA using LU103 oligonucleotides as sense and antisense primers to obtain LU103 amplicon, and (c) detecting LU103 amplicon in the test sample, where the LU103 oligonucleotides utilised in steps (a) and (b) have at least 50% sequence identity to the 269, 263, 225, 507 or 519 bp sequence and (3) a method detecting a target LU103 polynucleotide in a test sample suspended of containing the target, comprising: (a) contacting the test sample with at least 1 LU103 oligonucleotide as a sense primer and at least 1 LU103 oligonucleotide as an anti-sense primer and amplifying to obtain a first stage reaction product; (b) contacting the first stage reaction product with at least 1 other LU103 oligonucleotide to obtain a second stage reaction product, provided that the other LU103 oligonucleotide is located 3' to the LU103 oligonucleotides utilised in step (a) and is complementary to the first stage reaction product, and (c) detecting the second stage reaction product as an indication of the presence of the target LU103 polynucleotide, where the LU103 oligonucleotides utilised in steps (a) and (b) have at least 50% sequence identity to the 269, 263, 225, 507 or 519 bp sequence; (4) a purified polynucleotide derived from an LU103 gene, where the polynucleotide is capable of selectively hybridising to the nucleic acid of the LU103 gene and has at least 50% identity to the 269, 263, 225, 507 or 519 bp

sequence; (5) a recombinant expression system comprising a nucleic acid sequence that includes an ORF derived from LU103 operably linked to a control sequence compatible with a desired host, where the nucleic acid sequence has at least 50% identity to the 269, 263, 225, 507 or 519 bp sequence; (6) cell transfected with the recombinant expression system; (7) cell transfected with a nucleic acid sequence encoding at least 1 LU103 epitope, where the nucleic acid sequence has the 269, 263, 225, 507 or 519 bp sequence; (8) composition comprising a LU103 polynucleotide, where the polynucleotide has at least 50% identity to the 269, 263, 225, 507 or 519 bp sequence; (9) gene which encodes a LU103 protein which comprises an amino acid sequence with at least 50% identity with the 93 residue amino acid sequence given in the specification, and (10) gene comprising DNA having at least 50% identity with the 507 or 519 bp sequence.

USE - The methods and products of the invention may be used to detect, diagnose, stage, monitor, prognose, prevent, treat or determine the predisposition diseases and conditions of the lung, e.g. lung cancer.

Dwg.0/4

Burke
09/092296

(FILE 'CAPLUS' ENTERED AT 12:09:17 ON 25 OCT 1999)
L1 10217 SEA ABB=ON PLU=ON LUNG(3A)(DISEAS? OR DISORDER)
L2 583 SEA ABB=ON PLU=ON L1(S)(IDENTIF? OR DETECT? OR DET## OR DETERM? OR DIAGNOS?)
L3 14 SEA ABB=ON PLU=ON L2 AND (REAGENT OR EPITOPE OR LS147 OR LS 147)

L3 ANSWER 1 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1999:556186 CAPLUS
TITLE: Pigeon fanciers' **lung**: **identification** of **disease** -associated carbohydrate **epitopes** on pigeon intestinal mucin
AUTHOR(S): Baldwin, C. I.; Todd, A.; Bourke, S. J.; Allen, A.; Calvert, J. E.
CORPORATE SOURCE: Department of Immunology, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne, NE2 4HH, UK
SOURCE: Clin. Exp. Immunol. (1999), 117(2), 230-236
CODEN: CEXIAL; ISSN: 0009-9104
PUBLISHER: Blackwell Science Ltd.
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Pigeon intestinal mucin, a complex high mol. wt. glycoprotein, is a key antigen in the development of pigeon fanciers' lung (PFL). We have studied the specificity of antibodies to mucin in patients with PFL and asymptomatic antibody-pos. individuals. Extensive papain digestion, which removes the non-glycosylated regions of the mucin leaving the heavily glycosylated "bottle brush" regions, resulted in a 600-fold decrease in IgG3 antibody titers with little effect on IgG1 and IgG2 titers. This suggests that IgG1 and IgG2 are directed against the region rich in O-linked sugar chains while the majority of the IgG3 is directed against **epitopes** which are proteinase-sensitive. Lectin mapping of the carbohydrates present on pigeon intestinal mucin demonstrated high levels of exposed N-acetyl neuraminic acid, N-acetyl galactosamine and N-acetyl glucosamine, with lower levels of fucose and some galactose. Sera from pigeon fanciers inhibited binding of lectins specific for N-acetyl neuraminic acid, N-acetyl galactosamine, internal N-acetyl glucosamine and fucose. Sera from people with PFL, compared with sera from asymptomatic antibody-pos. fanciers, had significantly higher titers of antibody that inhibited binding of four lectins specific for N-acetyl galactosamine and one fucose-specific lectin, suggesting that these sugars may play a dominant role in disease-assocd. **epitopes**. The results suggest that different IgG subclasses recognize different **epitopes** on mucin and that the **epitopes** recognized by the major subclasses are present on the O-linked oligosaccharides. Further, the carbohydrate-specific anti-mucin antibodies produced by PFL

patients may differ in their specificity from those found in asymptomatic individuals.

L3 ANSWER 2 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1999:8152 CAPLUS
DOCUMENT NUMBER: 130:77055
TITLE: Protein LS170 and cDNA sequences useful for **detecting diseases** of the human **lung**
INVENTOR(S): Billing-Medel, Patricia A.; Cohen, Maurice; Colpitts, Tracey L.; Friedman, Paula N.; Gordon, Julian; Granados, Edward N.; Hodges, Steven C.; Klass, Michael R.; Kratochvil, Jon D.; Roberts-Rapp, Lisa; Russell, John C.; Stroupe, Stephen D.
PATENT ASSIGNEE(S): Abbott Laboratories, USA
SOURCE: PCT Int. Appl., 120 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9856951	A1	19981217	WO 1998-US11601	19980611
W: CA, JP				
RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE				
PRIORITY APPLN. INFO.:			US 1997-49183	19970611

AB A set of contiguous and partially overlapping cDNA sequences and polypeptides encoded thereby, designated as LS170 and transcribed from human lung tissue, is described. These sequences are useful for the detecting, diagnosing, staging, monitoring, prognosticating, in vivo imaging, preventing or treating, or detg. the predisposition of an individual to diseases and conditions of the lung, such as lung cancer. Also provided are antibodies which specifically bind to a LS170-encoded polypeptide or protein, and agonists or inhibitors which prevent action of tissue-specific LS170 polypeptides, which mols. are useful for the therapeutic treatment of lung diseases, tumors, or metastases.

L3 ANSWER 3 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1998:674032 CAPLUS
DOCUMENT NUMBER: 130:64487
TITLE: Immunohistochemical detection of multidrug resistance protein in human lung cancer and normal lung
AUTHOR(S): Wright, Scott R.; Boag, Alexander H.;
Searcher : Shears 308-4994

Valdimarsson, Gunnar; Hipfner, David R.; Campling, Barbara G.; Cole, Susan P. C.; Deeley, Roger G.
CORPORATE SOURCE: Departments of Pathology and Cancer Research Laboratories, Queen's University, Kingston, ON, K7L 3N6, Can.
SOURCE: Clin. Cancer Res. (1998), 4(9), 2279-2289
CODEN: CCREF4; ISSN: 1078-0432
PUBLISHER: American Association for Cancer Research
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Monoclonal antibody QCRL-1 is highly specific for a defined linear **epitope** in a relatively poorly conserved region of the human multidrug resistance protein (MRP). The authors have used QCRL-1 to examine MRP expression in archival and fresh snap-frozen samples of untreated small cell (SC) and non-small cell (NSC) lung cancers (LCs), as well as normal lung. The authors found that the majority (87%) of all histol. subtypes of NSCLC had detectable levels of MRP in most of the tumor mass. In a substantial proportion of adenocarcinomas (55%) and squamous cell carcinomas (28%), immunoreactivity approached that obtained with the highly multidrug resistant cell line H69AR from which the MRP was originally cloned. Both the level and frequency of MRP expression in untreated SCLC was significantly lower than in NSCLC. The MRP was detectable in only 56% of SCLC tumors and, in most cases, was expressed in small focal clusters of cells. Immunofluorescence studies of tumor tissue and normal lung confirmed the plasma membrane location of the MRP. However, in normal bronchial epithelium and seromucous glands, unlike in tumor cells, the MRP was detected only on basolateral membranes. In addn., strong MRP immunoreactivity was detected in reactive type II pneumocytes present in hyperplastic alveoli, but not in normal type I and type II pneumocytes. No potentially confounding correlation independent of its possible role in drug resistance was obsd. between MRP expression in untreated NSCLC and any clinicopathol. parameter examd., including overall survival.

L3 ANSWER 4 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1997:692333 CAPLUS
DOCUMENT NUMBER: 127:316564
TITLE: Diagnosis of chronic respiratory failure and pulmonary emphysema, and diagnostic **reagents**
INVENTOR(S): Okuda, Yukichi; Morita, Riichiro; Hanatani, Mitsuya; Matsuo, Katsuhiko
PATENT ASSIGNEE(S): Toa Gosei Chemical Industry Co., Ltd., Japan
SOURCE: Jpn. Kokai Tokkyo Koho, 4 pp.
CODEN: JKXXAF
DOCUMENT TYPE: Patent
LANGUAGE: Japanese
Searcher : Shears 308-4994

FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
JP 09274037	A2	19971021	JP 1996-104629	19960402

AB Detn. of human serum vascular endothelial growth factor/vascular permeability factor (VPF) is useful for accurate diagnosis of the title diseases. The detn. is carried out by using anti-VPF antibodies or VPF receptors. The serum level of VPF in patients with the diseases is much higher than healthy humans.

L3 ANSWER 5 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1997:44677 CAPLUS
DOCUMENT NUMBER: 126:55930
TITLE: Detection of neoplastic cells based on alternatively spliced transcripts of the p15INK4B and p16INK4A cyclin/CDK inhibitors
INVENTOR(S): Sidransky, David; Baylin, Stephen B.
PATENT ASSIGNEE(S): Johns Hopkins University School of Medicine, USA
SOURCE: PCT Int. Appl., 82 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9635704	A1	19961114	WO 1996-US6666	19960510
W: AU, CA, CN, JP, KR				
RW: AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE				
US 5767258	A	19980616	US 1995-439962	19950512
US 5856094	A	19990105	US 1995-497535	19950630
AU 9657399	A1	19961129	AU 1996-57399	19960510
PRIORITY APPLN. INFO.:			US 1995-439962	19950512
			US 1995-497535	19950630
			WO 1996-US6666	19960510

AB Novel cell cycle regulatory polynucleotide transcripts and their encoded polypeptides are provided which were identified as the products of alternatively spliced mRNA for cyclin/CDK inhibitors, p16INK4A and p15INK4B, and a 5' nucleotide sequence referred to as 5'ALT. The p16 and p15 genes colocalize to human chromosome 9p21, which has been identified as a region having homozygous deletions in many tumors. 5'ALT also resides on chromosome 9p21, just 5' of exon 2 of p15, and about 30 kb upstream from p16. Polynucleotide transcripts are provided in which a 5'ALT polynucleotide is operatively linked to (1) exon 2 and exon 3 of p15 or (2) exon 2 of

p15. These transcripts are assocd. with normal growth control and regulation of cellular proliferation and provide a means for the development of more accurate diagnostic, prognostic, and therapeutic regimes for disorders assocd. with control of cell cycle progression and cell differentiation and the loss of such control. Methylation of p16 DNA and a resultant decrease in p16 gene expression is assocd. with transcriptional block and is assocd. with a variety of neoplasms. Thus, a method for detecting a neoplasm in a subject by detecting methylation of 5'CpG islands in p16 DNA, or detecting p16 mRNA or polypeptide levels in a sample is also provided. Preferably, the method utilizes a methylation-sensitive restriction endonuclease in order to detect p16 methylation. The 5'ALT, or 5'ALT-p16 or 5'ALT-p15 polypeptides can also be used to produce antibodies which are immunoreactive or bind to **epitopes** of the resp. polypeptides.

L3 ANSWER 6 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1996:68700 CAPLUS
DOCUMENT NUMBER: 124:169878
TITLE: Difference in sero-diagnostic values among KL-6-associated mucins classified as Cluster 9
AUTHOR(S): Kohno, Nobuoki; Inoue, Yoshikazu; Hamada, Hironobu; Fujioka, Seiji; Fujino, Shun; Yokoyama, Akihito; Hiwada, Kunio; Ueda, Norifumi; Akiyama, Mitoshi
CORPORATE SOURCE: School Medicine, Ehime University, Ehime, 791-02, Japan
SOURCE: Int. J. Cancer, Suppl. (1994), 8(Third International IASLC Workshop on Lung Tumor and Differentiation Antigens, 1993), 81-3
CODEN: IJSUEZ; ISSN: 0898-6924
DOCUMENT TYPE: Journal
LANGUAGE: English

AB KL-6 classified as Cluster 9 (MUC-1) is a circulating high-mol.-wt. mucin-like mol. Serum level of KL-6 was measured by a sandwich assay using KL-6 antibody as not only a catcher but also as a tracer. The authors established 2 addnl. monoclonal antibodies (MAbs), LISA 101 and EH-123, reacting with KL-6 **epitopes** different from the **epitope** recognized by KL-6 antibody. The KL-6-assocd. mucins detected by the sandwich assay using LISA 101 or EH-123 antibody as a catcher and KL-6 antibody as a tracer were designated as LISA 1-6 and CAM 123-6 resp. The diagnostic values as the serum markers of KL-6, LISA 1-6 and CAM 123-6 were evaluated measuring their levels in the same serum from healthy individuals and from patients with pulmonary, pancreatic and breast adenocarcinomas. KL-6 was increased abnormally at high rates of more than 50% in pancreatic cancer and in benign lung diseases, LISA 1-6 only in pancreatic cancer, and CAM 123-6 only in pulmonary adenocarcinoma. In benign lung diseases, however, LISA 1-6 and CAM

123-6 were increased abnormally at the rates of only 5.3% and 0% resp. These observations clearly indicate that LISA 1-6 and CAM 123-6 constitute a part of KL-6, but that they are superior to KL-6 as tumor markers for pancreatic cancer and for pulmonary adenocarcinoma resp., because of their much lower false-pos. rates.

L3 ANSWER 7 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1994:104481 CAPLUS
DOCUMENT NUMBER: 120:104481
TITLE: Monoclonal antibodies against farmer's lung antigens having specific binding to IgG antibodies
AUTHOR(S): Kumar, Anoopa; Elms, Nancy; Kurup, Viswanath P.
CORPORATE SOURCE: Dep. Med., Med. Coll. Wisconsin, V., Milwaukee, WI, USA
SOURCE: Int. Arch. Allergy Immunol. (1993), 102(1), 67-71
CODEN: IAAIEG; ISSN: 1018-2438
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Hypersensitivity pneumonitis resulting from environmental exposure to Saccharopolyspora rectivirgula (Micropolyspora faeni) among farmers has been well recognized. The diagnosis of the disease depends on demonstration of circulating antibodies against S. rectivirgula. However, dependable pure antigens are not available for serodiagnosis. The authors employed hybridoma technol. to obtain monoclonal antibodies against S. rectivirgula antigens. These monoclonal antibodies were employed to purify antigens through affinity chromatog. When tested in ELISA, high levels of antibodies were demonstrated against these antigens in farmer's lung patient sera compared to exposed but asymptomatic individuals from the same household. In Western blots, patient sera reacted with components of crude antigens with mol. masses of 28, 35, 60, 65 and 68 kD and 4 components above 100 kD, while the monoclonal antibodies reacted only with the 60-kD protein. These purified antigens can be used as reliable **reagents** in the specific **diagnosis** of farmer's **lung diseases**.

L3 ANSWER 8 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1993:467319 CAPLUS
DOCUMENT NUMBER: 119:67319
TITLE: Test **reagent** for hemoptysic sputum for diagnosis of lung cancer
INVENTOR(S): Qin, Dexing
PATENT ASSIGNEE(S): Chinese Academy of Medical Sciences, Tumour Hospital, Peop. Rep. China
SOURCE: Faming Zhuanli Shenqing Gongkai Shuomingshu, 4 pp.
CODEN: CNXXEV

DOCUMENT TYPE: Patent
LANGUAGE: Chinese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
CN 1069341	A	19930224	CN 1991-105479	19910814

AB The title **reagent** contg. alc., H2O2, glacial AcOH, and guaiac gum is used for **detecting** hidden blood in sputum for alerting a possible **lung** cancer **disease**. Using the test **reagent**, the incidence of lung cancer was 40 times higher in pos.-responding people than in neg.-responding people.

L3 ANSWER 9 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1993:444470 CAPLUS
DOCUMENT NUMBER: 119:44470
TITLE: Automated determination of crosslinked fibrin derivatives in plasma
AUTHOR(S): Elms, M. J.; Bundesen, P. G.; Rowbury, D.; Goodall, S.; Wakeham. N.; Rowell, J. A.; Hillyard, C. J.; Rylatt, D. B.
CORPORATE SOURCE: Pathol. Dep., R. Brisbane Hosp., Brisbane, Australia
SOURCE: Blood Coagulation Fibrinolysis (1993), 4(1), 159-64
CODEN: BLFIE7; ISSN: 0957-5235
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Automated assays for the measurement of crosslinked fibrin derivs. in plasma (XL-FbDP) were developed by using latex beads coated with anti-D-dimer monoclonal antibody (DD-3B6/22) for both the Cobas Fara Chem. Centrifugal and the Cobas Mira analyzers (Roche, Basle, Switzerland). The analyzers were programmed to mix plasma and latex **reagent** simultaneously and analyze absorbance changes over a 10-15 min period. Results were interpolated by the analyzer from a std. curve derived from a polymer of D-dimer. Both assays had high precision (<5% CV) for values between 100 and 1000 ng/mL and provided clear discrimination between normal samples and samples from patients suffering from the thrombotic diseases deep vein thrombosis/pulmonary embolism and disseminated intravascular coagulation. The results obtained for XL-FbDP detn. with both methods compared well with established methods: a high correlation was obtained with a semiquant. manual latex method for both the Fara ($r = 0.92$) and Mira ($r = 0.83$) and correlations (r) of 0.81 (Fara) and 0.84 (Mira) were obtained with an EIA. Correlation between the 2 automated procedures was high ($r = 0.96$). The automated method will enable labs. to provide a rapid and accurate quantitation of

XL-FbDP.

L3 ANSWER 10 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1992:608482 CAPLUS
DOCUMENT NUMBER: 117:208482
TITLE: Detection of early platelet activation and prediagnosis of thrombotic events by immunoassay for platelet surface thrombospondin (TSP)
INVENTOR(S): Aiken, Martha L.; Painter, Richard G.
PATENT ASSIGNEE(S): University of Texas System, USA
SOURCE: PCT Int. Appl., 105 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9215886	A1	19920917	WO 1992-US1757	19920309
W: AT, AU, BB, BG, BR, CA, CH, CS, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, PL, RO, RU, SD, SE				
RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, DE, DK, ES, FR, GA, GB, GN, GR, IT, LU, MC, ML, MR, NL, SE, SN, TD, TG				
US 5256538	A	19931026	US 1991-668272	19910308
AU 9216554	A1	19921006	AU 1992-16554	19920309
PRIORITY APPLN. INFO.:			US 1991-668272	19910308
			WO 1992-US1757	19920309

AB Persons at risk for a thrombotic event are identified by early immunol. detn. of elevated platelet surface TSP using a (labeled) monoclonal antibody to TSP. Thus, IgG-coated magnetic beads were exposed to an anti-TSP monoclonal antibody and then mixed with paraformaldehyde-fixed human platelets. After magnetic sepn. of the beads, the no. of platelets remaining in suspension was inversely related to TSP surface expression on the platelets. Diagnostic kits contg. **reagents** for the assay are described.

L3 ANSWER 11 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1992:233421 CAPLUS
DOCUMENT NUMBER: 116:233421
TITLE: Human surfactant protein-A contains blood group A antigenic determinants
AUTHOR(S): Stahlman, Mildred T.; Gray, Mary E.; Ross, Gary F.; Hull, William M.; Wikenheiser, Kathryn; Dingle, Sharon; Zelenski-Low, Kay R.; Whitsett, Jeffrey A.
CORPORATE SOURCE: Sch. Med., Vanderbilt Univ., Nashville, TN, 37232-2370, USA
SOURCE: Pediatr. Res. (1992), 31(4, Pt. 1), 364-71

CODEN: PEREBL; ISSN: 0031-3998
DOCUMENT TYPE: Journal
LANGUAGE: English
AB A major blood group antigenic **epitope** was identified on human pulmonary surfactant protein A (SP-A). Monoclonal (MAb) and polyclonal antibodies generated against purified human SP-A aggregated blood group A human erythrocytes and immunostained epithelial cells in a variety of human tissues, consistent with the tissue distribution of major blood group antigens. SP-A MAb (MAb-8) agglutinated red cells and immunostained tissues from A or AB blood groups, but did not react with cells or tissues from O or B individuals. MAb-8 immunostaining of tissue from blood group A individuals was ablated by incubation with blood group A red cells. MAb and polyclonal antibodies directed against A blood group antigens reacted strongly with purified SP-A obtained from a blood group A individual with alveolar proteinosis. MAb and polyclonal antibodies specific for B blood group antigen failed to react with SP-A from this patient or from patients who were in blood group B. Reactivity of anti-blood group MAb was lost after treatment of SP-A with endoglycosidase-F, demonstrating its reactivity with an **epitope** dependent on the asparagine-linked oligosaccharide at asparagine 187. Reactivity of MAb-8 with SP-A persisted after endoglycosidase-F treatment, but was lost after digestion with collagenase as assessed by Western blot after SDS-PAGE. Reactivity of MAb to SP-A was sensitive to .beta.-elimination, supporting the presence of another blood group antigenic site distinct from the **epitope** dependent on the asparagine-linked carbohydrate. The finding that the SP-A mol. contains a major blood group **epitope** has implication for the clin. use of surfactant replacement prepns. and diagnostic **reagents** based on this protein.

L3 ANSWER 12 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1989:210563 CAPLUS
DOCUMENT NUMBER: 110:210563
TITLE: Monoclonal antibodies to angiotensin-converting enzyme: a powerful tool for lung and vessel studies
AUTHOR(S): Danilov, S.; Sakharov, I.; Martynov, A.; Faerman, A.; Muzykantov, V.; Klibanov, A.; Trakht, I.
CORPORATE SOURCE: Inst. Exp. Cardiol., Cardiol. Res. Cent., Moscow, 121552, USSR
SOURCE: J. Mol. Cell. Cardiol. (1989), 21(Suppl. 1), 165-70
CODEN: JMCDAY; ISSN: 0022-2828
DOCUMENT TYPE: Journal
LANGUAGE: English
AB A series (12 clones) of hybridomas were obtained, which produce

monoclonal antibodies (Mab) to 5 different **epitopes** of the angiotensin-converting enzyme (ACE) mol. These antibodies may be used to (1) map antigenic structure of ACE, including the study of immunol. heterogeneity of ACE from different organs and tissues; (2) study the immunohistochem. distribution of ACE in human tissues, including the diagnosis of sarcoidosis; (3) develop an ACE immunoassay, and (4) prep. an immunosorbent for large-scale ACE isolation and for ACE-apheresis. One of the antibodies, 9B9, when injected into the circulation of rat and monkey, accumulated with high specificity in the lungs as compared with either normal mouse IgG or other organs and blood. The highly specific and nontoxic accumulation of Mab 9B9 suggests that it also may be used for .gamma. scintigraphy visualization of the pulmonary vascular bed, detection of lung injury, and as a vector for targeted drug delivery to the lung.

L3 ANSWER 13 OF 14 CAPLUS COPYRIGHT 1999 ACS
ACCESSION NUMBER: 1988:585013 CAPLUS
DOCUMENT NUMBER: 109:185013
TITLE: Diisocyanate antigens that detect specific antibodies in exposed workers and guinea pigs
AUTHOR(S): Jin, Ruzhi; Karol, Meryl H.
CORPORATE SOURCE: Jilin Prov. Inst. Ind. Health Occup. Dis., Jilin, Peop. Rep. China
SOURCE: Chem. Res. Toxicol. (1988), 1(5), 288-93
CODEN: CRTOEC
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Evaluation of the immunol. contribution to the pathogenesis of isocyanate **lung disease** necessitates prepn. of isocyanate-protein conjugates to **detect** anti-isocyanate antibodies. Sera were obtained from 2 guinea pigs immunized with MDI and from 3 workers with occupational exposure to MDI. By use of Western blot anal., guinea pig IgG antibodies were best detected by the monomeric component of MDI-guinea pig serum albumin. ELISA addnl. indicated that conjugates which contained a high d. of hapten detected greater amts. of antibody than did conjugates contg. low amts. of hapten. The same procedures were then used to assess the amt. of MDI-specific IgG and IgE antibody in sera from symptomatic workers. Effective MDI-HSA antigens were those that were monomeric and had high haptenic content. Highly substituted conjugates of monoisocyanates (Ph isocyanate and p-tolyl isocyanate) with serum albumins were also effective in detecting antibodies to MDI. These results indicate the importance of the compn. of isocyanate conjugate antigens in detecting antibodies to diisocyanates and suggest that stds. be developed for prepn. and characterization of these diagnostic serol. **reagents**.

L3 ANSWER 14 OF 14 CAPLUS COPYRIGHT 1999 ACS

ACCESSION NUMBER: 1974:57087 CAPLUS
DOCUMENT NUMBER: 80:57087
TITLE: Modified procedure for evaluation of the lecithin/sphingomyelin ratio in amniotic fluid
AUTHOR(S): Coch, Emily; Meyer, John S.; Goldman, Gordon; Kessler, Gerald
CORPORATE SOURCE: Dep. Pathol. Lab. Med., Jew. Hosp., St. Louis, Mo., USA
SOURCE: Clin. Chem. (1973), 19(9), 967-72
CODEN: CLCHAU
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Several modifications were made to the procedure of L. Gluck, et al. (1971) for evaluation of amniotic fluid lecithin/sphingomyelin (L/S) ratios. The acetone-pptn. step was eliminated, resulting in a faster extn. procedure and higher concns. of phospholipids. For faster thin-layer chromatog., silica gel-impregnated sheets of glass-fiber and a modified solvent system are used. For spot detection, a noncorrosive Bi subnitrate **reagent** specific for lecithin and sphingomyelin and an I vapor method to confirm the L/S ratio were used. The Bi spray also detected significant specimen contamination by phospholipids from plasma or erythrocytes. The L/S ratio was evaluated by visually comparing the relative size and color intensity of the lecithin and sphingomyelin TLC spots. An L/S ratio of 5 or more is consistent with mature pulmonary function and an L/S ratio of less than 2 suggests fetal pulmonary immaturity.

(FILE 'MEDLINE, BIOSIS, EMBASE, LIFESCI, WPIDS, CONFSCI, SCISEARCH, JICST-EPLUS, PROMT, CANCERLIT' ENTERED AT 12:12:56 ON 25 OCT 1999)
L4 210 S L3
L5 113 DUP REM L4 (97 DUPLICATES REMOVED)
L6 14 S L5 AND (HYBRIDIZ? OR HYBRIDIS?)

L6 ANSWER 1 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1999-418749 [35] WPIDS
CROSS REFERENCE: 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-506364 [42]; 1998-520811 [44]; 1998-609887 [42]; 1999-059865 [05]; 1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-430031 [36]
DOC. NO. CPI: C1999-123038
TITLE: New isolated human genes encoding secreted polypeptides.
DERWENT CLASS: B04 D16
INVENTOR(S): CARTER, K C; DUAN, R D; FENG, P; FERRIE, A M; FLORENCE, C; FLORENCE, K; GREENE, J M; JANAT, F; KYAW, H; MOORE, P A; NI, J; ROSEN, C A; RUBEN, S M; SHI, Y; SOPPET, D R; WEI, Y; YU, G

PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 82
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9931117	A1	19990624	(199935)*	EN	536

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM HR HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9931117	A1	WO 1998-US27059	19981217

PRIORITY APPLN. INFO: US 1997-68369 19971219; US 1997-68006 19971218; US 1997-68007 19971218; US 1997-68008 19971218; US 1997-68053 19971218; US 1997-68054 19971218; US 1997-68057 19971218; US 1997-68064 19971218; US 1997-70923 19971218; US 1997-68169 19971219; US 1997-68365 19971219; US 1997-68367 19971219; US 1997-68368 19971219

AN 1999-418749 [35] WPIDS

CR 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-506364 [42]; 1998-520811 [44]; 1998-609887 [42]; 1999-059865 [05]; 1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-430031 [36]

AB WO 9931117 A UPAB: 19990908
NOVELTY - Isolated human nucleic acids (I) encoding secreted proteins, are new.

DETAILED DESCRIPTION - (I) comprises a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to:

(a) a PN fragment of one of a total of 110 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z which is **hybridizable** to one of the 110 defined cDNA sequences;

(b) a PN which is an (allelic) variant of one of the 110 defined cDNA sequences;

(c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 110

defined amino acid sequences given in the specification or a polypeptide fragment, domain or **epitope**, respectively, encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridizable** to one of the defined cDNA sequences;

(d) a PN which encodes a species homolog of one of the 110 defined polypeptides; or

(e) a PN capable of **hybridizing** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridize** under stringent conditions to a sequence of only A residues or of only T residues.

ATCC Deposit No. Z refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the ATCC. The deposit numbers are ATCC 209463, 209511, 209551.

INDEPENDENT CLAIMS are also included for the following:

(1) a recombinant vector comprising (I);

(2) a method of making a recombinant host cell comprising (I);

(3) a recombinant host cell produced by a method as in (2);

(4) an isolated polypeptide comprising an amino acid sequence at least 95% identical to a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 110 defined amino acid sequences or the encoded sequence included in ATCC Deposit No. Z;

(5) an isolated antibody that binds specifically to an isolated polypeptide as in (4);

(6) a recombinant host cell that expresses an isolated polypeptide as above;

(7) a method of making an isolated polypeptide by culturing the host cell of (6);

(8) the polypeptide produced by the method of (7);

(9) a gene corresponding to a cDNA sequence of the 110 defined amino acid sequences;

(10) a method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in (I); and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of the mutation;

(11) a method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of (4) in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide;

(12) a method for identifying a binding partner to the polypeptide of (4) comprising:

(a) contacting the polypeptide of (4) with a binding partner;
(b) determining whether the binding partner effects an activity of the polypeptide; and
(13) a method of identifying an activity in a biological assay, where the method comprises:
(a) expressing one of the 110 defined cDNAs in a cell;
(b) isolating the supernatant;
(c) detecting an activity in a biological assay; and
(d) identifying the protein in the supernatant having the activity; and
(14) the product produced by the method of (13).
ACTIVITY - Cytostatic; Nootropic; Neuroprotective; Osteopathic; Antiseborreic; Dermatological; Antipsoritic; Antidiabetic; Antiasthmatic.
MECHANISM OF ACTION - None given.
USE - The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 110 PNs, based on which tissues they are most highly expressed in, and include developing products for the diagnosis or treatment of cancer, tumors, neurodegenerative disorders, developmental abnormalities and fetal deficiencies, reproductive disorders, blood disorders, leukemias, diseases of the immune system, autoimmune diseases, hepatic and renal disease, lymphomas, inflammation, allergies, Alzheimer's and cognitive disorders, schizophrenia, disorders involving osteoclasts such as osteoporosis, arthritis, sepsis, acne, asthma, psoriasis, stroke, trauma, diseases of testes, lung or prostate, digestive/endocrine disorders, diabetes and AIDS. The polypeptides are also useful for identifying their binding partners (claimed).
Dwg.0/0

L6 ANSWER 2 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1999-303069 [25] WPIDS
CROSS REFERENCE: 1999-312869 [25]
DOC. NO. NON-CPI: N1999-227015
DOC. NO. CPI: C1999-089015
TITLE: New isolated human genes and the secreted polypeptides they encode.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): BREWER, L A; CARTER, K C; DUAN, D R; EBNER, R; ENDRESS, G A; FENG, P; FLORENCE, C; FLORENCE, K A; GREENE, J M; JANAT, F; KAYW, H; LAFLEUR, D W; MOORE, P A; NI, J; OLSEN, H S; ROSEN, C A; RUBEN, S M; SHI, Y; SOPPET, D R; WEI, Y; YOUNG, P
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC

COUNTRY COUNT: 83
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9922243	A1	19990506	(199925)*	EN	546

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GD GE GH GM HR HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW

AU 9912734 A 19990517 (199939)

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9922243	A1	WO 1998-US22376	19981023
AU 9912734	A	AU 1999-12734	19981023

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9912734	A	Based on	WO 9922243

PRIORITY APPLN. INFO: US 1997-63387 19971024; US 1997-62784 19971024; US 1997-63088 19971024; US 1997-63089 19971024; US 1997-63090 19971024; US 1997-63091 19971024; US 1997-63092 19971024; US 1997-63097 19971024; US 1997-63098 19971024; US 1997-63099 19971024; US 1997-63100 19971024; US 1997-63101 19971024; US 1997-63109 19971024; US 1997-63110 19971024; US 1997-63111 19971024; US 1997-63148 19971024; US 1997-63386 19971024

AN 1999-303069 [25] WPIDS
CR 1999-312869 [25]
AB WO 9922243 A UPAB: 19990707
NOVELTY - One hundred and forty eight isolated human genes and secreted proteins they encode are new.

DETAILED DESCRIPTION - An isolated nucleic acid molecule (NAM) (I) comprising a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to:

(a) a PN fragment of one of a total of 148 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z which is

hybridisable to one of the 148 defined cDNA sequence;

(b) a PN which is an (allelic) variant of one of the 148 defined cDNA sequences;

(c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 148 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the defined cDNA sequences;

(d) a PN which encodes a species homologue of one of the 148 defined polypeptides; or

(e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues.

INDEPENDENT CLAIMS are also included for:

(1) a recombinant vector comprising (I);

(2) a method of making a recombinant host cell comprising (I);

(3) a recombinant host cell produced by a method as in (2);

(4) an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 148 defined amino acid sequences or the encoded sequence included in ATCC Deposit No. Z;

(5) an isolated antibody that binds specifically to an isolated polypeptide as in (4); and

(6) a recombinant host cell that expresses a gene corresponding to a cDNA sequence of the 148 defined amino acid sequences.

Note: From the disclosure 'ATCC Deposit No. Z' refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the ATCC. The deposit numbers are: ATCC 209299, 209346, 209300, 209324.an isolated polypeptide as above;

ACTIVITY - None given.

MECHANISM OF ACTION - None given.

USE - The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also pathological conditions can be **diagnosed** by **determining** the amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 148 PNs, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and fetal deficiencies, blood disorders, leukemias, diseases of the immune system, autoimmune diseases, hepatic and renal disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's and cognitive

disorders, schizophrenia, prostate diseases, obesity, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, **diseases** of testes, **lung** or thymus, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for **identifying** their binding partners (claimed).

L6 ANSWER 3 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1999-167452 [14] WPIDS
DOC. NO. CPI: C1999-048990
TITLE: New isolated human genes encoding secreted polypeptides - useful for diagnosis and treatment of pathalogical diseases.
DERWENT CLASS: B04 D16
INVENTOR(S): BREWER, L A; EBNER, R; FERRIE, A M; GREENE, J M; JANAT, F; NI, J; OLSEN, H S; ROSEN, C A; RUBEN, S M; SOPPET, D R; YOUNG, P E; YU, G
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 82
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9907891	A1	19990218	(199914)*	EN	329
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW					
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM HR HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW					
AU 9887684	A	19990301	(199928)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9907891	A1	WO 1998-US16235	19980804
AU 9887684	A	AU 1998-87684	19980804

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9887684	A	Based on	WO 9907891

PRIORITY APPLN. INFO: US 1997-56732 19970819; US 1997-54798 19970805; US 1997-54803 19970805; US 1997-54804 19970805; US 1997-54806 19970805; US 1997-54807 19970805; US

Searcher : Shears 308-4994

1997-54808 19970805; US 1997-54809
19970805; US 1997-55309 19970805; US
1997-55310 19970805; US 1997-55311
19970805; US 1997-55312 19970805; US
1997-55386 19970805; US 1997-55970
19970818; US 1997-55986 19970818; US
1997-56364 19970819; US 1997-56365
19970819; US 1997-56366 19970819; US
1997-56367 19970819; US 1997-56370
19970819; US 1997-56371 19970819; US
1997-56557 19970819; US 1997-56563
19970819; US 1997-56731 19970819

AN 1999-167452 [14] WPIDS

AB WO 9907891 A UPAB: 19990412

NOVELTY - A total of 90 new isolated human genes encode secreted polypeptides which can be used in the diagnosis and treatment of pathalogical diseases such as cancers, neurological disorders, immune diseases, inflammation or blood disorders.

DETAILED DESCRIPTION - IDEPENDENT CLAIMS are included for the following: (1) An isolated nucleic acid molecule (NAM) (I) is claimed comprising a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to: (a) a PN fragment of one of a total of 90 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the 90 defined cDNA sequence; (b) a PN which is an (allelic) variant of one of the 90 defined cDNA sequences; (c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 90 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the defined cDNA sequences; (d) a PN which encodes a species homologue of one of the 90 defined polypeptides; or (e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues; (2) a recombinant vector comprising (I); (3) a method of making a recombinant host cell comprising (I); (4) a recombinant host cell produced by a method as in (3); (5) an isolated polypeptide comprising an amino acid sequence homologous to a sequence selected from a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 90 defined amino acid sequences or the encoded sequence included in ATCC Deposit No. Z; (6) an isolated antibody that binds specifically to an isolated polypeptide as in (5); (7) a recombinant host cell that expresses an isolated polypeptide as above; and (8) a gene corresponding to a cDNA sequence of the 90 defined amino acid

sequences. Note: From the disclosure 'ATCC Deposit No. Z' refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the ATCC. The deposit numbers are: ATCC 209146, 209177, 209179 and 209180.

USE - The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also pathological conditions can be **diagnosed** by **determining** the amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 90 PNs, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and foetal deficiencies, blood disorders, CNS disorders, diseases of the immune system, autoimmune diseases, hepatic and renal disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's and cognitive disorders, schizophrenia, cardiovascular disorders, prostate diseases, asthma, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, **diseases** of testes, **lung** or thymus, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for **identifying** their binding partners (claimed).

Dwg.0/0

L6 ANSWER 4 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD

ACCESSION NUMBER: 1999-153691 [13] WPIDS
DOC. NO. CPI: C1999-045418
TITLE: New isolated human genes and the secreted polypeptides they encode - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.
DERWENT CLASS: B04 D16
INVENTOR(S): CARTER, K C; ENDRESS, G A; FAN, P; FENG, P; KYAW, H; LAFLEUR, D W; LI, Y; MOORE, P A; ROSEN, C A; RUBEN, S M; SHI, Y; WEI, Y; ZENG, Z
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 82
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9906423	A1	19990211	(199913)*	EN	311

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM HR HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW

Searcher : Shears 308-4994

AU 9887634 A 19990222 (199927)

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9906423	A1	WO 1998-US15949	19980729
AU 9887634	A	AU 1998-87634	19980729

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9887634	A	Based on	WO 9906423

PRIORITY APPLN. INFO: US 1997-56730 19970819; US 1997-54209 19970730; US 1997-54211 19970730; US 1997-54212 19970730; US 1997-54213 19970730; US 1997-54214 19970730; US 1997-54215 19970730; US 1997-54217 19970730; US 1997-54218 19970730; US 1997-54234 19970730; US 1997-54236 19970730; US 1997-55968 19970818; US 1997-55969 19970818; US 1997-55972 19970818; US 1997-56534 19970819; US 1997-56543 19970819; US 1997-56554 19970819; US 1997-56561 19970819; US 1997-56727 19970819; US 1997-56729 19970819

AN 1999-153691 [13] WPIDS

AB WO 9906423 A UPAB: 19990331

An isolated nucleic acid molecule (NAM) (I) is claimed comprising a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to: (a) a PN fragment of one of a total of 83 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the 83 defined cDNA sequence; (b) a PN which is an (allelic) variant of one of the 83 defined cDNA sequences; (c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 83 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the defined cDNA sequences; (d) a PN which encodes a species homologue of one of the 83 defined polypeptides; or (e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues. Also claimed are: (1) a recombinant vector comprising (I); (2) a method of making a recombinant host cell

comprising (I); (3) a recombinant host cell produced by a method as in (2); (4) an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 83 defined amino acid sequences or the encoded sequence included in ATCC Deposit No. Z; (5) an isolated antibody that binds specifically to an isolated polypeptide as in (4); (6) a recombinant host cell that expresses an isolated polypeptide as above; and (7) a gene corresponding to a cDNA sequence of the 83 defined amino acid sequences. Note: From the disclosure ATCC Deposit No. Z refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the ATCC. The deposit numbers are: ATCC 209145, 209148, 209147.

USE- The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also pathological conditions can be **diagnosed** by **determining** the amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 83 PNs, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and foetal deficiencies, blood disorders, leukemias, diseases of the immune system, autoimmune diseases, hepatic and renal disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's and cognitive disorders, schizophrenia, restenosis, prostate diseases, obesity, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, **diseases** of **lung** or skin, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for **identifying** their binding partners (claimed).

Dwg.0/0

L6 ANSWER 5 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD

ACCESSION NUMBER: 1999-120770 [10] WPIDS

CROSS REFERENCE: 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-506364 [42]; 1998-520811 [44]; 1998-609887 [42]; 1999-059865 [05]; 1999-080881 [07]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-418749 [32]; 1999-430031 [36]

DOC. NO. CPI: C1999-035369

TITLE: New isolated human genes and the secreted polypeptides they encode - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.

Searcher : Shears 308-4994

DERWENT CLASS: B04 D16
INVENTOR(S): BREWER, L A; EBNER, R; FISCHER, C L; KYAW, H; LAFLEUR, D W; LI, Y; MOORE, P A; OLSEN, H S; ROSEN, C A; RUBEN, S M; SHI, Y; SOPPET, D R; ZENG, Z
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 83
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9902546	A1	19990121	(199910)*	EN	462
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW					
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM GW HR HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW					
AU 9884743	A	19990208	(199924)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9902546	A1	WO 1998-US13684	19980707
AU 9884743	A	AU 1998-84743	19980707

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9884743	A	Based on	WO 9902546

PRIORITY APPLN. INFO: US 1997-58785 19970912; US 1997-51916 19970708; US 1997-51918 19970708; US 1997-51919 19970708; US 1997-51920 19970708; US 1997-51925 19970708; US 1997-51926 19970708; US 1997-51928 19970708; US 1997-51929 19970708; US 1997-51930 19970708; US 1997-51931 19970708; US 1997-51932 19970708; US 1997-52732 19970708; US 1997-52733 19970708; US 1997-52793 19970708; US 1997-52795 19970708; US 1997-52803 19970708; US 1997-55684 19970818; US 1997-55722 19970818; US 1997-55723 19970818; US 1997-55947 19970818; US 1997-55948 19970818; US 1997-55949 19970818; US 1997-55950 19970818; US 1997-55953 19970818; US 1997-55954

Searcher : Shears 308-4994

19970818; US 1997-55964 19970818; US 1997-55984 19970818; US 1997-56360 19970818; US 1997-58660 19970912; US 1997-58661 19970912; US 1997-58664 19970912

AN 1999-120770 [10] WPIDS

CR 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-506364 [42]; 1998-520811 [44]; 1998-609887 [42]; 1999-059865 [05]; 1999-080881 [07]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-418749 [32]; 1999-430031 [36]

AB WO 9902546 A UPAB: 19990908

An isolated nucleic acid molecule (NAM) (I) is claimed comprising a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to: (a) a PN fragment of one of a total of 123 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the 123 defined cDNA sequence; (b) a PN which is an (allelic) variant of one of the 123 defined cDNA sequences; (c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 123 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the defined cDNA sequences; (d) a PN which encodes a species homologue of one of the 123 defined polypeptides; or (e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues. Also claimed are: (1) a recombinant vector comprising (I); (2) a method of making a recombinant host cell comprising (I); (3) a recombinant host cell produced by a method as in (2); (4) an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 123 defined amino acid sequences or the encoded sequence included in ATCC Deposit No. Z; (5) an isolated antibody that binds specifically to an isolated polypeptide as in (4); (6) a recombinant host cell that expresses an isolated polypeptide as above; and (7) a gene corresponding to a cDNA sequence of the 123 defined amino acid sequences. Note: From the disclosure 'ATCC Deposit No. Z' refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the ATCC. The deposit numbers are: ATCC 209119, 209124, 209125, 209126.

USE - The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also pathological conditions can be **diagnosed** by **determining** the

amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 123 PNs, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and foetal deficiencies, blood disorders, leukemias, diseases of the immune system, autoimmune diseases, hepatic and renal disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's and cognitive disorders, schizophrenia, restenosis, prostate diseases, obesity, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, **diseases** of testes, **lung** or thyroid, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for **identifying** their binding partners (claimed).
Dwg.0/0

L6 ANSWER 6 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1999-070066 [06] WPIDS
DOC. NO. CPI: C1999-020600
TITLE: New isolated human genes and the secreted polypeptides they encode - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.
DERWENT CLASS: B04 D16
INVENTOR(S): BREWER, L A; DUAN, R; EBNER, R; FERRIE, A M; FLORENCE, K A; GREENE, J M; HU, J; LAFLEUR, D W; MOORE, P A; NI, J; OLSEN, H S; ROSEN, C A; RUBEN, S M; SHI, Y; YOUNG, P
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 81
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9842738	A1	19981001	(199906)*	EN	385
RW: AT BE CH DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW					
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW					
AU 9865646	A	19981020	(199909)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9842738	A1	WO 1998-US5311	19980319

Searcher : Shears 308-4994

AU 9865646 A AU 1998-65646 19980319

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9865646	A	Based on	WO 9842738

PRIORITY APPLN. INFO: US 1997-50937 19970530; US 1997-41276 19970321; US 1997-41277 19970321; US 1997-41281 19970321; US 1997-42344 19970321; US 1997-48069 19970530; US 1997-48094 19970530; US 1997-48095 19970530; US 1997-48096 19970530; US 1997-48099 19970530; US 1997-48131 19970530; US 1997-48135 19970530; US 1997-48154 19970530; US 1997-48160 19970530; US 1997-48186 19970530; US 1997-48187 19970530; US 1997-48188 19970530; US 1997-48351 19970530; US 1997-48352 19970530; US 1997-48355 19970530

AN 1999-070066 [06] WPIDS

AB WO 9842738 A UPAB: 19990210

An isolated nucleic acid molecule (NAM) (I) comprises a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to:(a) a PN fragment of one of a total of 87 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the 87 defined cDNA sequences; (b) a PN which is an (allelic) variant of one of the 87 defined cDNA sequences; (c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 87 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the defined cDNA sequence; (d) a PN which encodes a species homologue of one of the 87 defined polypeptides; or (e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues. Also claimed are: (1) a recombinant vector comprising (I); (2) a method of making a recombinant host cell comprising (I); (3) a recombinant host cell produced by a method as in (2); (4) an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 87 defined amino acid sequences or the encoded sequence included in ATCC

Deposit No. Z; (5) an isolated antibody that binds specifically to an isolated polypeptide as in (4); (6) a recombinant host cell that expresses an isolated polypeptide as in (4); and (7) a gene corresponding to a cDNA sequence of the 87 defined amino acid sequences. Note: From the disclosure 'ATCC Deposit No. Z' refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the ATCC. The deposit numbers are: ATCC 97923, 97924, 97957, 97958, 209071, 209641, 209072 and 209073.

USE - The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also pathological conditions can be **diagnosed** by **determining** the amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 87 PNs, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and foetal deficiencies, blood disorders, leukaemias, diseases of the immune system (including allergies or asthma), lymphocytic diseases, brain associated diseases, hepatic and renal disease, lymphomas, inflammation, ischemic shock, Alzheimer's and cognitive disorders, schizophrenia, restenosis, prostate diseases, obesity, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, **diseases** of testis, **lung** or thymus, digestive/endocrine disorders, including metabolic regulation, malabsorption, gastritis and neoplasms, and AIDS. The polypeptides are also useful for **identifying** their binding partners (claimed).

Dwg.0/0

L6 ANSWER 7 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD

ACCESSION NUMBER: 1999-060335 [05] WPIDS
DOC. NO. NON-CPI: N1999-044750
DOC. NO. CPI: C1999-018022
TITLE: New LS170 nucleic acid from lung tissue - useful for **detecting**, monitoring, preventing and treating **lung disease**, particularly cancer.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): BILLING-MEDEL, P A; COHEN, M; COLPITTS, T L; FRIEDMAN, P N; GORDON, J; GRANADOS, E N; HODGES, S C; KLASS, M R; KRATOCHVIL, J D; ROBERTS-RAPP, L; RUSSELL, J C; STROUPE, S D
PATENT ASSIGNEE(S): (ABBO) ABBOTT LAB
COUNTRY COUNT: 20
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9856951	A1	19981217	(199905)*	EN	119

RW: AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE
W: CA JP

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9856951	A1	WO 1998-US11601	19980611

PRIORITY APPLN. INFO: US 1997-49183 19970611
AN 1999-060335 [05] WPIDS
AB WO 9856951 A UPAB: 19990210
Detection of a target LS170 polynucleotide (I) comprises treating a test sample with at least one LS170-specific nucleic acid (II) that has at least 50% identify with any of 9 sequences ((1)-(9), reproduced, of 251, 243, 258, 202, 458, 273, 200,1009 and 1027 bp, respectively), their fragments or complements. Also new are (a) LS170 sequences, or their fragments, that **hybridise** selectively to the LS170 gene and have at least 50% identity with (i) (1)-(9), other than (5) or their complements, or (ii) fragments of (1)-(7); (b) recombinant expression system containing (II) plus a control sequence; (c) cells transformed with this expression system; (d) LS170 polypeptides (III) at least 50% identical with 9 sequences ((23)-(31), of 256, 18, 26, 29, 12, 15, 11 or 19 amino acids (aa), reproduced), or their fragments, containing at least one LS170 **epitope**; (e) antibodies (Ab) specific for at least one LS170 **epitope** present on (III); (f) cells transformed with any of (1)-(9), their fragments or complements, encoding at least one **epitope** of LS170; (g) (I) having at least 50% identity with (i) any of (1)-(9), their fragments or complements or (ii) fragments of (1)-(7); (h) genes, or their fragments, encoding LS170 proteins at least 50% identical with (23); (j) genes, or their fragments, at least 50% identical with (8) or (9).

USE - LS170 represents a set of contiguous, partially overlapping sequences transcribed from lung tissue. They are used for **diagnosis**, staging, monitoring, in vivo imaging, prevention and treatment of **lung disease**, specifically cancer, and to indicate predisposition to such disease. Particularly **detection** of (I), LS170 antigens (using Ab in immunoassays) or anti-LS170 antibodies (using LS170 as antigen) is indicative of disease. Cells of (c) are used to express recombinant (III), used to raise Ab and for drug screening. LS170-related nucleic acid can be used to isolate related sequences; as standards and **reagents** in assays; as targets for drug screening, and as components or targets for therapy, e.g. as antisense, ribozyme or

triplex-forming agents. Ab can be used to deliver therapeutic agents to LS170-expressing cells; directly as therapeutic agents (by neutralising LS170 polypeptides); in competitive binding drug screens, and to generate anti-idiotypic antibodies for use in rational drug design.
Dwg.0/4

L6 ANSWER 8 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1999-059865 [05] WPIDS
CROSS REFERENCE: 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-506364 [42]; 1998-520811 [44]; 1998-609887 [42]; 1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-418749 [32]; 1999-430031 [36]
DOC. NO. NON-CPI: N1999-044486
DOC. NO. CPI: C1999-017701
TITLE: New isolated human genes and the secreted polypeptides they encode - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): BREWER, L A; CARTER, K C; DILLON, P J; EBNER, R; ENDRESS, G A; FAN, P; FENG, P; FERRIE, A M; FISCHER, C L; FLORENCE, C; FLORENCE, K; GREENE, J M; HU, J; KYAW, H; LAFLEUR, D W; LI, Y; MOORE, P A; NI, J; OLSEN, H S; ROSEN, C A; RUBEN, S M; SHI, Y; SOPPET, D R; WEI, Y; YOUNG, P; YU, G; ZENG, Z
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 82
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9854963	A2	19981210	(199905)*	EN	770
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW					
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW					
AU 9878120	A	19981221	(199919)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9854963	A2	WO 1998-US11422	19980604
AU 9878120	A	AU 1998-78120	19980604

Searcher : Shears 308-4994

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9878120	A	Based on	WO 9854963

PRIORITY APPLN. INFO: US 1997-70923 19971218; US 1997-48875 19970606; US 1997-48876 19970606; US 1997-48877 19970606; US 1997-48878 19970606; US 1997-48880 19970606; US 1997-48881 19970606; US 1997-48882 19970606; US 1997-48883 19970606; US 1997-48884 19970606; US 1997-48885 19970606; US 1997-48892 19970606; US 1997-48893 19970606; US 1997-48894 19970606; US 1997-48895 19970606; US 1997-48896 19970606; US 1997-48897 19970606; US 1997-48898 19970606; US 1997-48899 19970606; US 1997-48900 19970606; US 1997-48901 19970606; US 1997-48915 19970606; US 1997-48916 19970606; US 1997-48917 19970606; US 1997-48949 19970606; US 1997-48962 19970606; US 1997-48963 19970606; US 1997-48964 19970606; US 1997-48970 19970606; US 1997-48971 19970606; US 1997-48972 19970606; US 1997-48974 19970606; US 1997-49019 19970606; US 1997-49020 19970606; US 1997-49373 19970606; US 1997-49374 19970606; US 1997-49375 19970606; US 1997-57584 19970905; US 1997-57627 19970905; US 1997-57628 19970905; US 1997-57629 19970905; US 1997-57634 19970905; US 1997-57635 19970905; US 1997-57642 19970905; US 1997-57643 19970905; US 1997-57644 19970905; US 1997-57645 19970905; US 1997-57646 19970905; US 1997-57647 19970905; US 1997-57648 19970905; US 1997-57649 19970905; US 1997-57650 19970905; US 1997-57651 19970905; US 1997-57654 19970905; US 1997-57661 19970905; US 1997-57662 19970905; US 1997-57666 19970905; US 1997-57667 19970905; US 1997-57668 19970905; US 1997-57760 19970905; US 1997-57761 19970905; US 1997-57762 19970905; US 1997-57763 19970905; US

Searcher : Shears 308-4994

1997-57764 19970905; US 1997-57765 19970905; US 1997-57769 19970905; US 1997-57770 19970905; US 1997-57771 19970905; US 1997-57774 19970905; US 1997-57775 19970905; US 1997-57776 19970905; US 1997-57777 19970905; US 1997-57778 19970905

AN 1999-059865 [05] WPIDS

CR 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-506364 [42]; 1998-520811 [44]; 1998-609887 [42]; 1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-418749 [32]; 1999-430031 [36]

AB WO 9854963 A UPAB: 19990908

An isolated nucleic acid molecule (NAM) (I) comprising a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to: (a) a PN fragment of one of a total of 207 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the 207 defined cDNA sequence; (b) a PN which is an (allelic) variant of one of the 207 defined cDNA sequences; (c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 207 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the defined cDNA sequences; (d) a PN which encodes a species homologue of one of the 207 defined polypeptides; or (e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues. Also claimed are: (1) a recombinant vector comprising (I); (2) a method of making a recombinant host cell comprising (I); (3) a recombinant host cell produced by a method as in (2); (4) an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 207 defined amino acid sequences or the encoded sequence included in ATCC Deposit No. Z; (5) an isolated antibody that binds specifically to an isolated polypeptide as in (4); (6) a recombinant host cell that expresses an isolated polypeptide as above; and (7) a gene corresponding to a cDNA sequence of the 207 defined amino acid sequences.

Note: From the disclosure 'ATCC Deposit No. Z' refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the ATCC. The deposit numbers are: ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008,

209009, 209010, 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511,.

USE The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also pathological conditions can be **diagnosed** by **determining** the amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 207 PNs, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and foetal deficiencies, blood disorders, leukemias, diseases of the immune system, autoimmune diseases, hepatic and renal disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's and cognitive disorders, schizophrenia, restenosis, prostate diseases, obesity, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, **diseases** of testes, **lung** or thymus, digestive/endocrine disorders, infections and AIDS. The polypeptides are also useful for **identifying** their binding partners (claimed).

Dwg.0/0

L6 ANSWER 9 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD

ACCESSION NUMBER: 1998-609887 WPIDS

CROSS REFERENCE: 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-506364 [42]; 1998-520811 [44]; 1999-059865 [05]; 1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-418749 [32]; 1999-430031 [36]

DOC. NO. NON-CPI: N1998-474475

DOC. NO. CPI: C1998-182720

TITLE: New isolated human genes and the secreted polypeptides they encode - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.

DERWENT CLASS: B04 D16 S03

INVENTOR(S): BEDNARIK, D P; BREWER, L A; CARTER, K C; DUAN, R; EBNER, R; ENDRESS, G A; FENG, P; FERRIE, A M; FISCHER, C L; FLORENCE, K A; GREENE, J M; HU, J; NI, J; OLSEN, H S; ROSEN, C A; RUBEN, S M; SOPPET, D R; YOUNG, P E; YU, G; KYAW, H; LAFLEUR, D W; LI, Y; MOORE, P A; SHI, Y; ZENG, Z

PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC

COUNTRY COUNT: 81

PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG

WO 9839446 A2 19980911 (199851)* EN 447
RW: AT BE CH DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW
NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI
GB GE GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT
LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL
TJ TM TR TT UA UG US UZ VN YU ZW
AU 9865452 A 19980922 (199908)

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9839446	A2	WO 1998-US4482	19980306
AU 9865452	A	AU 1998-65452	19980306

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9865452	A	Based on	WO 9839446

PRIORITY APPLN. INFO: US 1997-57761 19970905; US 1997-38621
19970307; US 1997-40161 19970307; US
1997-40162 19970307; US 1997-40163
19970307; US 1997-40333 19970307; US
1997-40334 19970307; US 1997-40336
19970307; US 1997-40626 19970307; US
1997-43311 19970411; US 1997-43312
19970411; US 1997-43313 19970411; US
1997-43314 19970411; US 1997-43315
19970411; US 1997-43568 19970411; US
1997-43569 19970411; US 1997-43576
19970411; US 1997-43578 19970411; US
1997-43580 19970411; US 1997-43669
19970411; US 1997-43670 19970411; US
1997-43671 19970411; US 1997-43672
19970411; US 1997-43674 19970411; US
1997-47492 19970523; US 1997-47500
19970523; US 1997-47501 19970523; US
1997-47502 19970523; US 1997-47503
19970523; US 1997-47581 19970523; US
1997-47582 19970523; US 1997-47583
19970523; US 1997-47584 19970523; US
1997-47585 19970523; US 1997-47586
19970523; US 1997-47587 19970523; US
1997-47588 19970523; US 1997-47589
19970523; US 1997-47590 19970523; US

Searcher : Shears 308-4994

1997-47592 19970523; US 1997-47593
19970523; US 1997-47594 19970523; US
1997-47595 19970523; US 1997-47596
19970523; US 1997-47597 19970523; US
1997-47598 19970523; US 1997-47599
19970523; US 1997-47600 19970523; US
1997-47601 19970523; US 1997-47612
19970523; US 1997-47613 19970523; US
1997-47614 19970523; US 1997-47615
19970523; US 1997-47617 19970523; US
1997-47618 19970523; US 1997-47632
19970523; US 1997-47633 19970523; US
1997-48964 19970606; US 1997-48974
19970606; US 1997-56630 19970822; US
1997-56631 19970822; US 1997-56632
19970822; US 1997-56636 19970822; US
1997-56637 19970822; US 1997-56662
19970822; US 1997-56664 19970822; US
1997-56845 19970822; US 1997-56862
19970822; US 1997-56864 19970822; US
1997-56872 19970822; US 1997-56874
19970822; US 1997-56875 19970822; US
1997-56876 19970822; US 1997-56877
19970822; US 1997-56878 19970822; US
1997-56879 19970822; US 1997-56880
19970822; US 1997-56881 19970822; US
1997-56882 19970822; US 1997-56884
19970822; US 1997-56886 19970822; US
1997-56887 19970822; US 1997-56888
19970822; US 1997-56889 19970822; US
1997-56892 19970822; US 1997-56893
19970822; US 1997-56894 19970822; US
1997-56903 19970822; US 1997-56908
19970822; US 1997-56909 19970822; US
1997-56910 19970822; US 1997-56911
19970822; US 1997-57650 19970905

AN 1998-609887 WPIDS

CR 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-506364 [42]; 1998-520811 [44]; 1999-059865 [05]; 1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-418749 [32]; 1999-430031 [36]

AB WO 9839446 A UPAB: 19990908

An isolated nucleic acid molecule (NAM) (I) comprising a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to:

(a) a PN fragment of one of a total of 70 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z, which is

hybridisable to one of the 70 defined cDNA sequence;

(b) a PN which is an (allelic) variant of one of the 70 defined cDNA sequences;

(c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 70 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the defined cDNA sequence;

(d) a PN which encodes a species homologue of one of the 70 defined polypeptides., or

(e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues.

Also claimed are:

(1) a recombinant vector comprising (I),

(2) a method of making a recombinant host cell comprising (1);

(3) a recombinant host cell produced by a method as in (2):

(4) an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 70 defined amino acid sequences or the encoded sequence included in ATCC Deposit No. Z;

(5) an isolated antibody that binds specifically to an isolated polypeptide as in (4);

(6) a recombinant host cell that expresses an isolated polypeptide as in (4), and

(7) a gene corresponding to a cDNA sequence of the 70 defined amino acid sequences.

Note., From the disclosure 'ATCC Deposit No. Z' refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the ATCC. The deposit numbers are- ATCC 97900, 97901, 97903, 97898, 209010, 97897, 97899, 97897, 97976, 97904, 97926, 209043-209050, 209085, 209084, 209236.

USE - The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also pathological conditions can be **diagnosed** by **determining** the amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 70 PNs, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and foetal deficiencies, blood disorders, leukaemias, diseases of the immune system (including allergies or asthma), lymphocytic diseases, brain associated diseases, hepatic and renal disease, lymphomas,

inflammation, ischemic shock, Alzheimer's and cognitive disorders, schizophrenia, restenosis, prostate diseases, obesity, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, **diseases** of testis, **lung** or thymus, digestive/endocrine disorders, including metabolic regulation, malabsorption, gastritis and neoplasms, and AIDS. The polypeptides are also useful for **identifying** their binding partners (claimed).
Dwg.0/10

L6 ANSWER 10 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-594496 [50] WPIDS
DOC. NO. NON-CPI: N1998-462620
DOC. NO. CPI: C1998-178291
TITLE: New isolated human genes and secreted polypeptide(s) they encode - useful for the diagnosis and treatment of e.g. cancers, CNS disorders, immune system disorders, inflammatory disease and bacterial infections.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): FENG, P; NI, J; ROSEN, C A; RUBEN, S M; YU, G
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 82
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9845712	A2	19981015	(199850)*	EN	141

RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW

AU 9869529	A	19981030	(199911)

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9845712	A2	WO 1998-US6801	19980407
AU 9869529	A	AU 1998-69529	19980407

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9869529	A	Based on	WO 9845712

PRIORITY APPLN. INFO: US 1997-48184 19970530; US 1997-42726 19970408; US 1997-42727 19970408; US 1997-42728 19970408; US 1997-42754 19970408; US 1997-42825 19970408; US 1997-48068 19970530; US 1997-48070 19970530

AN 1998-594496 [50] WPIDS

AB WO 9845712 A UPAB: 19990210

An isolated nucleic acid molecule (NAM) (I) is claimed comprising a polynucleotide (PN) having a nucleotide sequence (NS) at least 95% identical to:(a) a PN fragment of a total of 20 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the 20 defined cDNA sequences; (b) a PN which is an (allelic) variant of one of the 20 defined cDNA sequence; (c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 20 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z which is **hybridisable** to one of the defined cDNA sequences; (d) a PN which encodes a species homologue of one of the 20 defined polypeptides; or (e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues. Also claimed are: (1) a recombinant vector comprising (I); (2) a method of making a recombinant host cell comprising (I); (3) a recombinant host cell produced by a method as in (2); (4) an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 20 defined amino acid sequences or the encoded sequence included in ATCC Deposit No. Z; (5) an isolated antibody that binds specifically to an isolated polypeptide as in (4); (6) a recombinant host cell that expresses an isolated polypeptide as in (4); (7) a gene corresponding to a cDNA sequence of one of the 20 nucleotide sequences given in the specification. Note: From the disclosure ''ATCC deposit No. Z'' refers to ATCC No's 97955 and 209074.

USE - The PNs and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein of gene therapy. Also pathological conditions can be **diagnosed** by **determining** the amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new PNs (claimed). Specific uses are described for each of the 20 PNs, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of central nervous system (CNS) and immune system diseases, reproductive disorders, cancers,

congenital malformations, degenerative diseases, trauma, inflammatory disease, neoplasia, metabolic disorders, diseases in testes, placenta, liver, brain and activated T cells, spleen **diseases**, **lung diseases**, heart **diseases**, rhabdomyosarcoma and disorders of the endocrine system or other endocrinopathies, e.g. endocrine polyglandular syndrome, endocrinoma, and endocrine ophthalmopathy, osteoclastoma and other bone remodelling disorders, bacterial infections and sepsis. The polypeptides are also useful for **identifying** their binding partners (claimed).
Dwg.0/0

L6 ANSWER 11 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-506364 [43] WPIDS
CROSS REFERENCE: 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-520811 [44]; 1998-609887 [42]; 1999-059865 [05]; 1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-418749 [32]; 1999-430031 [36]
DOC. NO. NON-CPI: N1998-394741
DOC. NO. CPI: C1998-152795
TITLE: New isolated human genes and the secreted polypeptide(s) they encode - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): BEDNARIK, D P; BREWER, L A; CARTER, K C; DUAN, R; EBNER, R; ENDRESS, G A; FENG, P; FERRIE, A M; FISCHER, C L; FLORENCE, K A; GREENE, J M; HU, J; NI, J; OLSEN, H S; ROSEN, C A; RUBEN, S M; SOPPET, D R; YOUNG, P E; YU, G; KYAW, H; LAFLEUR, D W; LI, Y; MOORE, P A; SHI, Y; ZENG, Z
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 81
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9839448	A2	19980911	(199843)*	EN	721
RW: AT BE CH DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW					
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW					
AU 9865453	A	19980922	(199908)		
AU 9891304	A	19990322	(199931)		

Searcher : Shears 308-4994

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9839448	A2	WO 1998-US4493	19980306
AU 9865453	A	AU 1998-65453	19980306
AU 9891304	A	AU 1998-91304	19980903

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9865453	A	Based on	WO 9839448
AU 9891304	A	Based on	WO 9911293

PRIORITY APPLN. INFO: US 1997-61060 19971002; US 1997-38621 19970307; US 1997-40161 19970307; US 1997-40162 19970307; US 1997-40163 19970307; US 1997-40333 19970307

AN 1998-506364 [43] WPIDS

CR 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35]; 1998-427559 [35]; 1998-520811 [44]; 1998-609887 [42]; 1999-059865 [05]; 1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11]; 1999-204988 [17]; 1999-418749 [32]; 1999-430031 [36]

AB WO 9839448 A UPAB: 19990908

An isolated nucleic acid molecule (I) comprises a polynucleotide (PN) having a nucleotide sequence at least 95% identical to: (a) a PN fragment of one of a total of 186 defined human cDNA sequences given in the specification or a PN fragment of the cDNA sequence included in ATCC Deposit No: Z, which is **hybridisable** to one of the 186 defined cDNA sequences; (b) a PN which is an (allelic) variant of one of the 186 defined cDNA sequences; (c) a PN encoding a biologically active polypeptide or a polypeptide fragment, domain or **epitope** of one of the 186 defined amino acid sequences given in the specification or a polypeptide fragment encoded by a cDNA sequence included in ATCC Deposit No. Z, which is **hybridisable** to one of the defined cDNA sequences; (d) a PN which encodes a species homologue of one of the 186 defined polypeptides; or (e) a PN capable of **hybridising** under stringent conditions to any one of the PNs specified in (a)-(d), where the PN does not **hybridise** under stringent conditions to a sequence of only A residues or of only T residues.

Also claimed is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from: (a) a polypeptide fragment (preferably having biological activity), domain, **epitope**, secreted form, full-length protein, (allelic) variant or species homologue of one of the 186 defined

amino acid sequences or the encoded sequence included in ATCC Deposit No: Z.

Note: From the disclosure ''ATCC Deposit No: Z'' refers to the representative clones, each containing a subset of the defined cDNA sequences, which have been deposited with the American Type Culture Collection. The deposit numbers are: ATCC 97897-97904 and ATCC 209076, 209215, 209139, 209011, 209119, 209235, 209627 and 209043-209050.

USE - The polynucleotides and their corresponding secreted polypeptides are useful for preventing, treating or ameliorating medical conditions (claimed), e.g. by protein or gene therapy. Also, pathological conditions can be **diagnosed** by **determining** the amount of the new polypeptides in a sample or by **determining** the presence of mutations in the new polynucleotides (claimed). Specific uses are described for each of the 186 polynucleotides, based on which tissues they are most highly expressed in, and include developing products for the **diagnosis** or treatment of: cancer; tumours; neurodegenerative disorders; developmental abnormalities and foetal deficiencies; blood disorders; leukaemias; diseases of the immune system (including allergies or asthma); lymphocytic diseases; brain associated diseases; hepatic and renal disease; lymphomas; inflammation; ischaemic shock; Alzheimer's and cognitive disorders; schizophrenia; restenosis; prostate diseases; obesity; disorders involving osteoclasts such as osteoporosis, arthritis or malignancies; **diseases** of testis, **lung** or thymus; thyroiditis and thyroid tumours; digestive/endocrine disorders, including metabolic regulation, malabsorption, gastritis and neoplasms; and AIDS. The polypeptides are also useful for **identifying** their binding partners (claimed).

Dwg.0/0

L6 ANSWER 12 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-437479 [37] WPIDS
DOC. NO. NON-CPI: N1998-340776
DOC. NO. CPI: C1998-133112
TITLE: New nucleic acid for the **lung** **disease** marker LU105 - polypeptides, antibodies and genes, used for **diagnosis**, prevention, treatment of **lung** **disease**, specifically cancer.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): BILLING-MEDEL, P A; COHEN, M; COLPITTS, T L; FRIEDMAN, P N; GORDON, J; GRANADOS, E N; HODGES, S C; KLASS, M R; KRATOCHVIL, J D; ROBERTSRAPP, L; RUSSELL, J C; STROUPE, S D
PATENT ASSIGNEE(S): (ABBO) ABBOTT LAB
COUNTRY COUNT: 19
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9833926	A1	19980806	(199837)*	EN	117

RW: AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE
W: CA JP

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9833926	A1	WO 1998-US1766	19980130

PRIORITY APPLN. INFO: US 1997-791710 19970131
AN 1998-437479 [37] WPIDS
AB WO 9833926 A UPAB: 19980916
A method for detecting target LU105 nucleic acid (I) comprises treating a sample with at least one LU105-specific nucleic acid (II), or its complement. (II) is at least 50% identical with 190, 244, 225, 114, 562 or 519 bp sequences given in the specification, or their fragments and complements.

Also claimed are:

(1) (I) or its fragments able to **hybridise** selectively to the LU105 gene and having at least 50% identity with the 190, 244, 225, 114, 562 or 519 bp sequences given above;

(2) recombinant expression systems including (I) and control sequence;

(3) cells transformed with this expression system;

(4) LU105 polypeptides (III) at least 50% identical with the 104, 26, 19, 21, 18 or 19 amino acid (aa) sequences given in the specification or their fragments;

(5) antibodies (Ab) that bind to at least one LU105 **epitope** present in (III);

(6) cells transformed with the 190, 244, 225, 114, 562 or 519 bp sequences described above;

(7) LU105 specific nucleic acid (II); and

(8) genes, or their fragments, that encode a protein at least 50% identical with the 104 aa sequences as in (4).

USE - LU105 is a **lung disease** marker. Cells as in (3) are used to express recombinant (III) which are used to raise Ab. Ab are used to **detect** the LU105 antigen, and correspondingly this antigen is used to **detect** specific antibodies, in usual immunoassays. (I) and (III) are used for **diagnosis**, staging, monitoring, prognosis, prevention, treatment (e.g. using antisense molecules, ribozymes, Ab or other antagonists) and **determination** of susceptibility to, **lung disease**, specifically cancer. (III) are also used to screen for specific binding agents, potentially useful

therapeutically. LU105 is a marker for **lung disease** (present at high concentration, in altered form or in an unusual body compartment).

ADVANTAGE - LU105 can be detected in blood, plasma or serum in an inexpensive, non-invasive test.

Dwg.0/6

L6 ANSWER 13 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-333327 [29] WPIDS
DOC. NO. NON-CPI: N1998-260122
DOC. NO. CPI: C1998-103375
TITLE: Human chemokine beta-13 polypeptide - useful in diagnosis and treatment of immune-system related disorders e.g. cancer of the immune system, leukaemias, autoimmune diseases etc..
DERWENT CLASS: B04 D16 S03
INVENTOR(S): LI, H; SEIBEL, G
PATENT ASSIGNEE(S): (HUMA-N) HUMAN GENOME SCI INC
COUNTRY COUNT: 79
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9824908	A1	19980611	(199829)*	EN	86

RW: AT BE CH DE DK EA ES FI FR GB GH GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW

AU 9853834	A	19980629	(199845)

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9824908	A1	WO 1997-US23144	19971205
AU 9853834	A	AU 1998-53834	19971205

FILING DETAILS:

PATENT NO	KIND		PATENT NO
AU 9853834	A	Based on	WO 9824908

PRIORITY APPLN. INFO: US 1996-32432 19961205
AN 1998-333327 [29] WPIDS
AB WO 9824908 A UPAB: 19980722
Isolated human chemokine beta -13 (CK beta -13) polypeptide

comprises amino acid sequence: (a) 95 % identical or more to 93 amino acid sequence (I)given in the specification, (b) mature CK beta -13 polypeptide sequence comprising amino acids 25-93 or 29-93 of sequence (I); (c) encoded by cDNA clone contained in ATCC 97113; or (d) a sequence complementary to the sequences as in (a)-(b).

Also claimed are: (1) polypeptide comprising **epitope** -bearing portion of CK beta -13 protein comprising amino acids Thr22-Gly28, Asn30-Leu47, Thr56-Val65 or Phe70-Trp83 of sequence (I); (2) nucleic acid molecules as follows: (I) encoding CK beta -13 polypeptide sequences as above; (ii) complementary to (I); (iii) **hybridising** to (I) but not to polynucleotide with sequence consisting of only A or only T residues; (iv) encoding **epitope**-bearing portion of (I), optionally as in (1); (v) having 282 bp sequence (II) and optionally encoding CK beta -13 polypeptide sequences as above; (vi) at least 95 % identical to sequence encoding polypeptide with amino acids n-93, 1-m and n-m (n and m are integers between 1-35 and 77-93 respectively) of (I); and (vii) encoding polypeptide which is a portion of CK beta -13 sequence encoded by cDNA clone in ATCC 97113 excluding 1-35 amino acids from amino terminus and/or 1-17 amino acids from carboxy terminus; (3) recombinant vectors produced by inserting (I) into vector; (4) host cells produced by introducing the vector as in (4); and (5) antibody binding specifically to CK beta -13 polypeptide.

USE - The polypeptides and nucleic acids are useful in **diagnosis** and treatment of immune system-related disorders in mammals (preferably humans). Such disorders include tumours, cancers, interstitial **lung disease** and disregulation of immune cell function including leukaemias, lymphomas, autoimmune diseases etc. For example, certain tissues in mammals with cancer of the immune system express enhanced/decreased levels of CK beta -13 and mRNA encoding CK beta -13, and **diagnosis** can be achieved by assaying CK beta -13 gene expression and comparing to standard levels. The polypeptides can be administered therapeutically in pharmaceutical compositions e.g. to treat immune system-related disorders as above, treat sepsis, inhibit bone marrow stem cell colony formation during cancer therapy, regulate haematopoiesis, stimulate wound healing etc. Compositions comprising the polynucleotides may also be administered, especially to express CK beta -13 polypeptide in hosts to treat dysfunctions associated with aberrant endogenous CK beta -13 activity. The polynucleotides are also useful for mapping of chromosomes/chromosome sites. The polypeptides are useful to screen for agonists and antagonists of CK beta -13 activity. The antibodies are useful **diagnostically** or therapeutically e.g. as antagonists to treat subjects requiring CK beta -13 reduction.
Dwg.0/4

L6 ANSWER 14 OF 14 WPIDS COPYRIGHT 1999 DERWENT INFORMATION LTD
ACCESSION NUMBER: 1998-286957 [25] WPIDS
Searcher : Shears 308-4994

DOC. NO. NON-CPI: N1998-225472
DOC. NO. CPI: C1998-088988
TITLE: Lung tissue derived polynucleotide LU103 - useful to **detect**, **diagnose**, stage, monitor, prognose, prevent, treat or **determine** pre-disposition to **lung disease**, e.g. **lung** cancer.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): COHEN, M; FRIEDMAN, P N; GORDON, J; HODGES, S C; KLASS, M R; KRATOCHVILL, J D; ROBERTS-RAPP, L; RUSSELL, J C; STROUPE, S D
PATENT ASSIGNEE(S): (ABBO) ABBOTT LAB
COUNTRY COUNT: 19
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9820143	A1	19980514	(199825)*	EN	86
RW: AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL PT SE					
W: CA JP					

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9820143	A1	WO 1997-US20680	19971105

PRIORITY APPLN. INFO: US 1996-744211 19961105
AN 1998-286957 [25] WPIDS
AB WO 9820143 A UPAB: 19980715
The following are claimed: (1) a method for detecting the presence of a target LU103 polynucleotide in a test sample, comprising: (a) contacting the sample with at least 1 LU103-specific polynucleotide, and (b) detecting the target LU103 polynucleotide in the test sample, where the LU103 polynucleotide has at least 50% identity to the 269, 263, 225, 507 or 519 bp nucleic acid sequence given in the specification; (2) a method for detecting LU103 mRNA in a test sample, comprising: (a) performing reverse transcription with at least 1 primer in order to produce cDNA; (b) amplifying the cDNA using LU103 oligonucleotides as sense and antisense primers to obtain LU103 amplicon, and (c) detecting LU103 amplicon in the test sample, where the LU103 oligonucleotides utilised in steps (a) and (b) have at least 50% sequence identity to the 269, 263, 225, 507 or 519 bp sequence and (3) a method detecting a target LU103 polynucleotide in a test sample suspended of containing the target, comprising: (a) contacting the test sample with at least 1 LU103 oligonucleotide as a sense primer and at least 1 LU103 oligonucleotide as an anti-sense primer and amplifying to obtain a

first stage reaction product; (b) contacting the first stage reaction product with at least 1 other LU103 oligonucleotide to obtain a second stage reaction product, provided that the other LU103 oligonucleotide is located 3' to the LU103 oligonucleotides utilised in step (a) and is complementary to the first stage reaction product, and (c) detecting the second stage reaction product as an indication of the presence of the target LU103 polynucleotide, where the LU103 oligonucleotides utilised in steps (a) and (b) have at least 50% sequence identity to the 269, 263, 225, 507 or 519 bp sequence; (4) a purified polynucleotide derived from an LU103 gene, where the polynucleotide is capable of selectively **hybridising** to the nucleic acid of the LU103 gene and has at least 50% identity to the 269, 263, 225, 507 or 519 bp sequence; (5) a recombinant expression system comprising a nucleic acid sequence that includes an ORF derived from LU103 operably linked to a control sequence compatible with a desired host, where the nucleic acid sequence has at least 50% identity to the 269, 263, 225, 507 or 519 bp sequence; (6) cell transfected with the recombinant expression system; (7) cell transfected with a nucleic acid sequence encoding at least 1 LU103 **epitope**, where the nucleic acid sequence has the 269, 263, 225, 507 or 519 bp sequence; (8) composition comprising a LU103 polynucleotide, where the polynucleotide has at least 50% identity to the 269, 263, 225, 507 or 519 bp sequence; (9) gene which encodes a LU103 protein which comprises an amino acid sequence with at least 50% identity with the 93 residue amino acid sequence given in the specification, and (10) gene comprising DNA having at least 50% identity with the 507 or 519 bp sequence.

USE - The methods and products of the invention may be used to detect, diagnose, stage, monitor, prognose, prevent, treat or determine the predisposition diseases and conditions of the lung, e.g. lung cancer.

Dwg.0/4

(FILE 'USPATFULL' ENTERED AT 12:18:19 ON 25 OCT 1999)

L9 7 S L2(S)(REAGENT OR EPITOPE OR LS147 OR LS 147)

L9 ANSWER 1 OF 7 USPATFULL

ACCESSION NUMBER: 1999:96216 USPATFULL

TITLE: **Reagents** and methods useful for **detecting diseases** of the **lung**

INVENTOR(S): Cohen, Maurice, Highland Park, IL, United States
Friedman, Paula N., Deerfield, IL, United States
Gordon, Julian, Lake Bluff, IL, United States
Hodges, Steven C., Buffalo Grove, IL, United States
Klass, Michael R., Libertyville, IL, United States

Searcher : Shears 308-4994

Kratochvil, Jon D., Kenosha, WI, United States
Roberts-Rapp, Lisa, Gurnee, IL, United States
Russell, John C., Kenosha, WI, United States
Stroupe, Steven D., Libertyville, IL, United States
PATENT ASSIGNEE(S): Abbott Laboratories, Abbott Park, IL, United States (U.S. corporation)

	NUMBER	DATE
PATENT INFORMATION:	US 5939265	19990817
APPLICATION INFO.:	US 1997-964725	19971105 (8)

RELATED APPLN. INFO.: Continuation-in-part of Ser. No. US 1996-744211, filed on 5 Nov 1996, now abandoned
DOCUMENT TYPE: Utility
PRIMARY EXAMINER: Degren, Nancy
ASSISTANT EXAMINER: Wang, Andrew
LEGAL REPRESENTATIVE: Becker, Cheryl L.; Goller, Mimi C.
NUMBER OF CLAIMS: 21
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 6 Drawing Figure(s); 6 Drawing Page(s)
LINE COUNT: 3052
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB A set of contiguous and partially overlapping RNA sequences and polypeptides encoded thereby, designated as LU103 and transcribed from lung tissue is described. A fully sequenced clone representing the longest continuous sequence of LU103 is also disclosed.. These sequences are useful for detecting, diagnosing, staging, monitoring, prognosticating, preventing or treating, or determining the predisposition of an individual to diseases and conditions of the lung such as lung cancer.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
INCL INCLM: 435/006.000
INCLS: 435/172.300; 435/320.100; 435/325.000; 536/023.100; 536/023.500
NCL NCLM: 435/006.000
NCLS: 435/320.100; 435/325.000; 536/023.100; 536/023.500

L9 ANSWER 2 OF 7 USPATFULL
ACCESSION NUMBER: 1999:43653 USPATFULL
TITLE: Amidinohydrazones as protease inhibitors

INVENTOR(S): Soll, Richard M., Lawrenceville, NJ, United States
Lu, Tianbao, Exton, PA, United States
Fedde, Cynthia L., Warrington, PA, United States
Tomczuk, Bruce E., Collegeville, PA, United States
Illig, Carl, Phoenixville, PA, United States

PATENT ASSIGNEE(S): 3-Dimensional Pharmaceuticals, Inc., Exton, PA, United States (U.S. corporation)

	NUMBER	DATE
PATENT INFORMATION:	US 5891909	19990406
APPLICATION INFO.:	US 1997-828160	19970327 (8)

	NUMBER	DATE
PRIORITY INFORMATION:	US 1996-14317	19960329 (60)

DOCUMENT TYPE: Utility
PRIMARY EXAMINER: Richter, Johann
ASSISTANT EXAMINER: Oswecki, Jane C.
LEGAL REPRESENTATIVE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
NUMBER OF CLAIMS: 76
EXEMPLARY CLAIM: 1
LINE COUNT: 5001
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB Amidino and benzamidino compounds, including compounds of the formula: ##STR1## wherein R.sup.1 -R.sup.4, R.sup.6 -R.sup.9, Y, Z, n and m are set forth in the specification, as well as hydrates, solvates or pharmaceutically acceptable salts thereof, that inhibit proteolytic enzymes such as thrombin are described. Also described are methods for preparing the compounds of Formula I.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

INCL INCLM: 514/517.000
INCLS: 514/309.000; 514/312.000; 514/345.000; 514/361.000; 514/401.000; 514/406.000; 514/518.000; 514/530.000; 514/567.000; 514/632.000; 546/141.000; 546/172.000; 546/290.000; 548/127.000; 548/353.100; 548/367.100; 549/230.000; 558/056.000; 560/303.000; 562/431.000; 562/432.000; 564/228.000

NCL NCLM: 514/517.000
NCLS: 514/309.000; 514/312.000; 514/345.000; 514/361.000; 514/401.000; 514/406.000; 514/518.000; 514/530.000; 514/567.000; 514/632.000; 546/141.000; 546/172.000; 546/290.000; 548/127.000; 548/353.100; 548/367.100; 549/230.000; 558/056.000; 560/303.000; 562/431.000; 562/432.000; 564/228.000

L9 ANSWER 3 OF 7 USPATFULL
ACCESSION NUMBER: 1998:157475 USPATFULL
TITLE: Hybridomas producing monoclonal antibodies to new mucin epitopes
INVENTOR(S): Linsley, Peter S., Seattle, WA, United States
Horn, Diane, Seattle, WA, United States
Searcher : Shears 308-4994

PATENT ASSIGNEE(S): Brown, Joseph P., Seattle, WA, United States
Sanofi, Paris, France (non-U.S. corporation)

	NUMBER	DATE
PATENT INFORMATION:	US 5849876	19981215
APPLICATION INFO.:	US 1994-179875	19940111 (8)

RELATED APPLN. INFO.: Continuation of Ser. No. US 1992-913740, filed on 14 Jul 1992, now abandoned which is a continuation of Ser. No. US 1987-104511, filed on 8 Oct 1987, now abandoned which is a continuation-in-part of Ser. No. US 1986-932781, filed on 19 Nov 1986, now abandoned
DOCUMENT TYPE: Utility
PRIMARY EXAMINER: Scheiner, Laurie
LEGAL REPRESENTATIVE: Merchant, Gould, Smith, Edell, Welter & Schmidt
NUMBER OF CLAIMS: 3
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 5 Drawing Figure(s); 5 Drawing Page(s)
LINE COUNT: 1849
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB Novel hybridoma cell lines producing monoclonal antibodies reactive with purified mucin antigens from normal and tumor sources are generated using mucins, including purified mucins from tumor sources. Epitopes on mucin antigens from normal and tumor souPPLICATION INFO.: US 1996-698401 19960815 (8)

RELATED APPLN. INFO.: Continuation-in-part of Ser. No. US 1995-536939, filed on 29 Sep 1995, now abandoned
DOCUMENT TYPE: Utility
PRIMARY EXAMINER: Ivy, C. Warren
ASSISTANT EXAMINER: Covington, Raymond
LEGAL REPRESENTATIVE: Sterne, Kessler Goldstein & Fox P.L.L.C.
NUMBER OF CLAIMS: 49
EXEMPLARY CLAIM: 1
LINE COUNT: 4363
CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB Compounds of the formula: ##STR1## wherein R.sup.1 --R.sup.4, R.sup.7 --R.sup.8, R.sup.a, R.sup.b, R.sup.c, Y, Z, n and m are set forth in the specification, as well as hydrates, solvates or pharmaceutically acceptable salts thereof, that inhibit a number of proteolytic enzymes are described. Also described are methods for preparing the compounds of Formula I.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
INCL INCLM: 514/255.000
INCLS: 514/317.000; 514/330.000; 514/331.000; 514/603.000; 514/604.000; 514/634.000; 514/822.000; 544/398.000; 544/399.000; 544/400.000; 544/402.000; 546/021.000; 546/229.000; 546/230.000; 546/231.000; 564/237.000
Searcher : Shears 308-4994

NCL NCLM: 514/255.000
NCLS: 514/317.000; 514/330.000; 514/331.000; 514/603.000; 514/604.000; 514/634.000; 514/822.000; 544/398.000; 544/399.000; 544/400.000; 544/402.000; 546/021.000; 546/229.000; 546/230.000; 546/231.000; 564/237.000

L9 ANSWER 5 OF 7 USPATFULL
ACCESSION NUMBER: 1998:69171 USPATFULL
TITLE: Cell cycle regulatory gene
INVENTOR(S): Sidransky, David, Baltimore, MD, United States
PATENT ASSIGNEE(S): The Johns Hopkins University School of Medicine, Baltimore, MD, United States (U.S. corporation)

	NUMBER	DATE
PATENT INFORMATION:	US 5767258	19980616
APPLICATION INFO.:	US 1995-439962	19950512 (8)

DOCUMENT TYPE: Utility
PRIMARY EXAMINER: Elliot, George C.
ASSISTANT EXAMINER: McGarry, Sean
LEGAL REPRESENTATIVE: Fish & Richardson, P.C.
NUMBER OF CLAIMS: 9
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 5 Drawing Figure(s); 3 Drawing Page(s)
LINE COUNT: 1506
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
AB A novel cell cycle regulatory gene called 5'ALT is disclosed. Methods for determining mutations or polymorphisms in 5'ALT or 5'ALT regulated genes in tissues are also provided. Novel 5'ALT-p16 and 5'ALT-p15 transcripts and truncated expression products are also described.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
INCL INCLM: 536/023.100
INCLS: 536/023.500; 435/006.000; 435/172.100; 435/320.100; 435/325.000; 435/243.000
NCL NCLM: 536/023.100
NCLS: 435/006.000; 435/243.000; 435/320.100; 435/325.000; 536/023.500

L9 ANSWER 6 OF 7 USPATFULL
ACCESSION NUMBER: 85:8964 USPATFULL
TITLE: .alpha.-Aminoboronic acid peptides
INVENTOR(S): Shenvi, Ashokkumar B., Wilmington, DE, United States
Kettner, Charles A., Wilmington, DE, United States
PATENT ASSIGNEE(S): E. I. Du Pont de Nemours and Company, Wilmington, DE, United States (U.S. corporation)
Searcher : Shears 308-4994

	NUMBER	DATE
PATENT INFORMATION:	US 4499082	19850212
APPLICATION INFO.:	US 1983-558362	19831205 (6)

DOCUMENT TYPE: Utility
PRIMARY EXAMINER: Phillips, Delbert R.
LEGAL REPRESENTATIVE: Hallquist, Scott G.
NUMBER OF CLAIMS: 19
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 3 Drawing Figure(s); 3 Drawing Page(s)
LINE COUNT: 1528
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
AB Peptides comprising C-terminal .alpha.-aminoboronic acid residues are potent, reversible inhibitors of proteolytic enzymes.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
INCL INCLM: 514/002.000
INCLS: 260/112.500R; 514/018.000; 514/019.000
NCL NCLM: 514/002.000
NCLS: 260/001.000; 514/018.000; 514/019.000; 514/506.000; 530/331.000; 562/007.000; 930/010.000; 930/023.000; 930/250.000

L9 ANSWER 7 OF 7 USPATFULL
ACCESSION NUMBER: 85:3359 USPATFULL
TITLE: Determination of oxidized .alpha.-1-proteinase inhibitor in serum or plasma
INVENTOR(S): Travis, James, Athens, GA, United States
PATENT ASSIGNEE(S): University of Georgia Research Foundation, Inc., Athens, GA, United States (U.S. corporation)

	NUMBER	DATE
PATENT INFORMATION:	US 4493891	19850115
APPLICATION INFO.:	US 1982-402442	19820727 (6)

DOCUMENT TYPE: Utility
PRIMARY EXAMINER: Kepplinger, Esther M.
LEGAL REPRESENTATIVE: Oblon, Fisher, Spivak, McClelland & Maier
NUMBER OF CLAIMS: 13
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 2 Drawing Figure(s); 1 Drawing Page(s)
LINE COUNT: 603
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
AB A new method of determining oxidized .alpha.-1-proteinase inhibitor in serum or plasma for use in studying the development of chronic obstructive lung disease is disclosed. Levels of oxidized .alpha.-1-proteinase inhibitor indicate the potential for emphysema development with higher levels showing a decrease in

lung protection against elastolytic enzymes such as elastase. This method can be used for patients with a potential for chronic obstructive lung disease rather than having to use bronchial lavage methods for such patients. No other method is known to exist for determining oxidized .alpha.-1-proteinase inhibitor in serum or plasma.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.
INCL INCLM: 435/023.000
INCLS: 435/184.000
NCL NCLM: 435/023.000
NCLS: 435/184.000

(FILE 'CAPLUS, MEDLINE, BIOSIS, EMBASE, LIFESCI, WPIDS, CONFSCI, SCISEARCH, JICST-EPLUS, PROMT, CANCERLIT, USPATFULL' ENTERED AT 12:20:18 ON 25 OCT 1999)

L10 134 SEA ABB=ON PLU=ON (BILLING MEDEL P? OR BILLING P? OR MEDEL P? OR MEDEL BILLING P?)/AU
L11 26222 SEA ABB=ON PLU=ON COHEN M?/AU
L12 123 SEA ABB=ON PLU=ON COLPITTS T?/AU
L13 2767 SEA ABB=ON PLU=ON FRIEDMAN P?/AU
L14 262 SEA ABB=ON PLU=ON KLASS M?/AU
L15 12902 SEA ABB=ON PLU=ON RUSSELL J?/AU
L16 310 SEA ABB=ON PLU=ON STROUPE S?/AU
L17 48 SEA ABB=ON PLU=ON L10 AND L11 AND L12 AND L13 AND L14 AND L15 AND L16
L18 51 SEA ABB=ON PLU=ON L10 AND (L11 OR L12 OR L13 OR L14 OR L15 OR L16)
L19 62 SEA ABB=ON PLU=ON L11 AND (L12 OR L13 OR L14 OR L15)
L20 51 SEA ABB=ON PLU=ON L12 AND (L13 OR L14 OR L15)
L21 62 SEA ABB=ON PLU=ON L13 AND (L14 OR L15)
L22 59 SEA ABB=ON PLU=ON L14 AND L15

L26 7 SEA ABB=ON PLU=ON (L17 OR L18 OR L19 OR L20 OR L21 OR L22) AND L1
L27 4 DUP REM L26 (3 DUPLICATES REMOVED)

Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
source 1. .311
/organism="Homo sapiens"
/note="Organ: brain; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
/db_xref="ATCC (inhost):124516"
/db_xref="taxon:9606"
/clone_lib="Cerebellum II"
/tissue_type="cerebellum"
/dev_stage="adult"

BASE COUNT 59 a 110 c 74 g 64 t 4 others
ORIGIN

Query Match 12.8%; Score 23; DB 11; Length 311;
Best Local Similarity 89.7%; Pred. No. 3.57e-06;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 8 TGGCCACTGGGGGGCCTGGGCTGCCCCTT 36
|||||||| |||| ||||||||||||||
Qy 12 TGGCCACTATGGGGTCTGGGCTGCCCCTT 40

RESULT 8
LOCUS AA323964 311 bp mRNA EST 20-APR-1997
DEFINITION EST26816 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA323964
NID g1976290
VERSION AA323964.1 GI:1976290
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT On Apr 14, 1993 this sequence version replaced gi:693635.

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA

RESULT 10
LOCUS AA323964 311 bp mRNA EST 20-APR-1997
DEFINITION EST26816 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA323964
NID g1976290
VERSION AA323964.1 GI:1976290
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,

Mon Oct 25 11:53:39 1999

Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT On Apr 14, 1993 this sequence version replaced gi:693635.

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .311
/organism="Homo sapiens"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
/db_xref="ATCC (inhost):124516"
/db_xref="taxon:9606"
/clone_lib="Cerebellum II"
/tissue_type="cerebellum"
/dev_stage="adult"
BASE COUNT 59 a 110 c 74 g 64 t 4 others
ORIGIN

Query Match 9.7%; Score 23; DB 11; Length 311;
Best Local Similarity 89.7%; Pred. No. 3.76e-05;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Db       8 TGGCCACTGGGGGGCCTGGGCTGCCCCTT 36
           |||||||| |||| ||||||||||||||
Qy      43 TGGCCACTATGGGGTCTGGGCTGCCCCTT 71
```

```
RESULT    9
ID   US-08-418-444A-1 STANDARD; DNA; UNC; 3088 BP.
AC   xxxxxx
DT
DE   Sequence 1, Application US/08418444A
CC   Sequence 1, Application US/08418444A
CC   Patent No. 5773688
CC    GENERAL INFORMATION:
CC      APPLICANT:  KURODA, HISAO
CC      APPLICANT:  HIROTA, NAOHIKO
CC      APPLICANT:  ITO, KAZUTOSHI
CC      TITLE OF INVENTION:  GENE EXPRESSION VECTOR USING THE GENE
CC      TITLE OF INVENTION:  EXPRESSION REGULATING REGION OF THE ADP RIBOSYLATION
CC      TITLE OF INVENTION:  FACTOR
CC      NUMBER OF SEQUENCES:  9
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
CC        STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC        CITY:  ARLINGTON
CC        STATE:  VIRGINIA
CC        COUNTRY:  USA
CC        ZIP:  22202
CC      COMPUTER READABLE FORM:
C         MEDIUM TYPE:  Floppy disk
C         COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/418,444A
CC        FILING DATE:  07-APR-1995
CC        CLASSIFICATION:  800
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  JP HEI 6-71048
CC        FILING DATE:  08-APR-1994
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Oblon, No. 5773688man F.
CC        REGISTRATION NUMBER:  24,618
CC        REFERENCE/DOCKET NUMBER:  2589-024-0
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703) 413-3000
CC        TELEFAX:  (703) 413-2220
CC        TELEX:  248855 OPAT UR
CC    INFORMATION FOR SEQ ID NO:  1:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  3088 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  double
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
```

```
SQ   SEQUENCE 3088 BP; 716 A; 761 C; 672 G; 939 T; 0 OTHER.

  Query Match             9.6%;  Score 22;  DB 3;  Length 3088;
  Best Local Similarity  72.0%;  Pred. No. 1.54e-01;
  Matches    36;  Conservative    0;  Mismatches  14;  Indels    0;  Gaps    0;

Db      15 CCGAAGATGGGCCAGAGGGGGGCCAGGGGCTGCCCAGGCGCCCTGCTGGC 64
           || | || || | ||||| || | |||||| |||||| | || |  ||||
Cp      90 CCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGC 41
```

```
CC       TELEX:  248855 OPAT UR
CC    INFORMATION FOR SEQ ID NO:  1:
CC      SEQUENCE CHARACTERISTICS:
CC        LENGTH:  3088 base pairs
CC        TYPE:  nucleic acid
CC        STRANDEDNESS:  double
CC        TOPOLOGY:  linear
CC      MOLECULE TYPE:  DNA (genomic)
SQ   SEQUENCE 3088 BP; 716 A; 761 C; 672 G; 939 T; 0 OTHER.

  Query Match           12.2%;  Score 22;  DB 3;  Length 3088;
  Best Local Similarity  72.0%;  Pred. No. 7.70e-02;
  Matches    36;  Conservative    0;  Mismatches  14;  Indels   0;  Gaps   0;

      15 CCGAAGATGGGCCAGAGGGGGGCCAGGGGCTGCCCAGGCGCCCTGCTGGC 64
         || | || || | ||||| || | |||||| |||||| | || |  ||||
      63 CCAAGGAGGGTCAAGAGGAGGACAAGGGGCAGCCCAGACCCCATAGTGGC 14
```

```
RESULT    8
ID   US-08-418-444A-1 STANDARD; DNA; UNC; 3088 BP.
AC   xxxxxx
DT
DE   Sequence 1, Application US/08418444A
CC   Sequence 1, Application US/08418444A
CC   Patent No. 5773688
CC    GENERAL INFORMATION:
CC      APPLICANT: 'KURODA, HISAO
CC      APPLICANT:  HIROTA, NAOHIKO
CC      APPLICANT:  ITO, KAZUTOSHI
CC      TITLE OF INVENTION:  GENE EXPRESSION VECTOR USING THE GENE
CC      TITLE OF INVENTION:  EXPRESSION REGULATING REGION OF THE ADP RIBOSYLATION
CC      TITLE OF INVENTION:  FACTOR
CC      NUMBER OF SEQUENCES:  9
CC      CORRESPONDENCE ADDRESS:
CC        ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
CC        STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CC        CITY:  ARLINGTON
CC        STATE:  VIRGINIA
CC        COUNTRY:  USA
CC        ZIP:  22202
CC      COMPUTER READABLE FORM:
CC        MEDIUM TYPE:  Floppy disk
CC        COMPUTER:  IBM PC compatible
CC        OPERATING SYSTEM:  PC-DOS/MS-DOS
CC        SOFTWARE:  PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC        APPLICATION NUMBER:  US/08/418,444A
CC        FILING DATE:  07-APR-1995
CC        CLASSIFICATION:  800
CC      PRIOR APPLICATION DATA:
CC        APPLICATION NUMBER:  JP HEI 6-71048
CC        FILING DATE:  08-APR-1994
CC      ATTORNEY/AGENT INFORMATION:
CC        NAME:  Oblon, No. 5773688man F.
CC        REGISTRATION NUMBER:  24,618
CC        REFERENCE/DOCKET NUMBER:  2589-024-0
CC      TELECOMMUNICATION INFORMATION:
CC        TELEPHONE:  (703) 413-3000
CC        TELEFAX:  (703) 413-2220
```